Exploring alternative oxidase (AOX) as a functional marker candidate for efficient somatic embryogenesis in Daucus carota L.

António Miguel da Costa Magriço e Frederico

Tese apresentada à Universidade de Évora
para obtenção do Grau de Doutor em Ciências Agrárias
Especialidade: Biotecnologia

ORIENTADOR (A/ES): *Professor Doutor Augusto António Vieira Peixe*
Professor Doutor Manuel Pedro Salema Fevereiro
Doutora Elena Baena-González

DECLARAÇÃO RELATIVA ÀS CONDIÇÕES DE REPRODUÇÃO DA TESE

Nome: António Miguel da Costa Magriço e Frederico
Endereço electrónico: amcmf@uevora.pt
Telefone: 00351 - 927572460
Cartão de Cidadão: 11161262 - 4ZY8

Título: Exploring alternative oxidase (AOX) as a functional marker candidate for efficient somatic embryogenesis in *Daucus carota* L.

Orientadores: Prof. Dr. Augusto Peixe; Prof. Dr. Pedro Fevereiro; Dra. Elena Baena-González

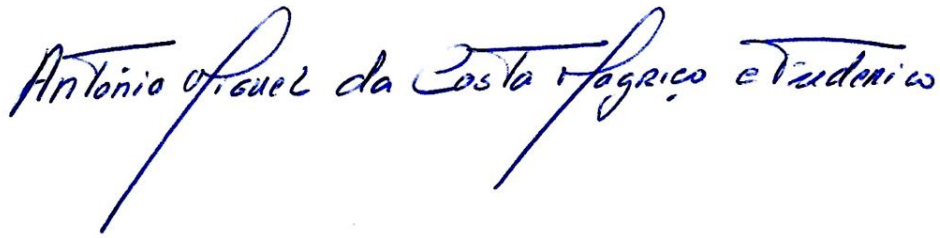
Ano de conclusão: 2017

Designação do Doutoramento: Doutoramento em Ciências Agrárias - Biotecnologia

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE. APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

IFFA/UE, 24/03/2017

Assinatura:



This work was funded by FEDER Funds through the Operational Program for Competitiveness Factors - COMPETE, the European Commission - Marie Curie Excellence Chairs (MEXC-CT-2004-006669), national funds through FCT - Foundation for Science and Technology under the Strategic Project PEst-C/AGR/UI0115/2011, Project: PTDC/EBB-BIO/099268/2008 and an individual fellowship to AMF - SFRH/BD/64987/2009.

…A hidden connection is stronger than an obvious one…

Heraclitus of Ephesus

535 BC - 475 BC

In memory of my ascendants:

António Magriço

Maria Jesus

António Frederico

Carmina Mendes

Manuel Frederico

Acknowledgements

At the final stage of this thesis, I could not leave without acknowledging the ones that directly or indirectly contributed to the final achievements and its conclusion. In particular, I would like to acknowledge to:

- My initial supervisors, Prof. Dr. Birgit Arnholdt-Schmitt and Dr. Hélia Cardoso, from whom I learned a lot. I deeply regret the situation that, in the end, would not allow us to pursue this project together. Nevertheless, I would like to express here my deepest acknowledgement to both;

- My current supervisors, Prof. Dr. Pedro Fevereiro, Dr. Elena Baena-González and Prof. Dr. Augusto Peixe, who embraced the challenge to carry through a thesis in the context of an uncommon situation. I deeply acknowledge your courage, words, and all the advices and help throughout the thesis writing process;

- The board of governance of the University of Évora with the support rendered during the complex process of supervision change;

- The IIFA and ICAAM former and present directors and staff, by the support and encouragement throughout the thesis development difficult process. I deeply acknowledge all your efforts;

- My former Molecular and Plant Breeding and Biotechnology Laboratories colleagues. Somehow, all of you contributed in a special way to the development of the present thesis;

- The ICAAM laboratories staff, without whose help, throughout the scientific work development and throughout the fun and party moments, the development of the present thesis would have been impossible;

- Virginia Sobral, a special person, with special feelings. Without whom, nothing of this would become reality;

- Henrique Nande, for all the help in the images management and by the encouraging words in the moments when the emotional strength vanished;

- My friends, without their help, I could never have achieved this goal;

- My brothers, Manuel and José, safety shoulders and examples of courage and strength;

- And finally, my mother, Maria, I deeply thank all your words, given in the good and bad moments. Somehow, they transformed me, and granted me the strength to undergo this complex task.

Resumo

Exploração da oxidase alternativa como marcador funcional inovador para a embriogénese somática eficaz em *Daucus carota* L..

A embriogénese somática é o mais conhecido exemplo de reprogramação celular. *Daucus carota* L. foi a primeira espécie onde a totipotência foi comprovada, através da embriogénese somática. No entanto, mesmo em *Daucus*, considerada como modelo e facilmente induzível, existe uma influência genética na capacidade das células para serem reprogramadas. Neste sentido, a identificação de marcadores para a 'fácil reprogramação' pode ajudar ao desenvolvimento de marcadores funcionais para a eficiente propagação *in vitro* de genótipos recalcitrantes. Tendo como objetivo explorar esta questão e usando a oxidase alternativa como o gene candidato a marcador funcional, 28 genótipos de *Daucus* foram induzidos a realizar embriogénese somática indireta. Desses, 25 responderam ao processo, produzindo 139 linhas celulares. A eficiência embriogénica das linhas foi avaliada utilizando um método de fenotipagem em dois passos, concebido para a seleção dos fenótipos estáveis. Após o primeiro passo de fenotipagem, 41 linhas com eficiência embriogénica estável foram avaliadas e caracterizadas em relação à diversidade genética utilizando o método cTBP. Destas, 22 foram selecionados para o segundo passo de fenotipagem. Finalmente, 8 linhas celulares foram identificadas, como altamente estáveis para as eficiências embriogénicas extremas. Quatro eram muito eficientes ou muito eficientes / eficientes, e as restantes não-embriogénicas. Estas foram estabelecidas como a coleção base utilizada para uma avaliação mais aprofundada, tanto em relação à ploidia, como na investigação da oxidase alternativa. Foi detectada poliploidia em três das linhas celulares, independentemente da sua capacidade embriogénica. No entanto, considerando a mais recente informação, em que se refere a poliploidia como uma importante fonte de variabilidade para a tolerância ao stress geral, a sua utilização no presente estudo foi considerada. Os três genes da oxidase alternativa foram amplificados a partir da coleção base e explorados tendo em vista a identificação de posições polimórficas associadas com os fenótipos definidos. No total, 290 sequências foram amplificadas, das quais 47 foram identificadas como polimórficas. Destas, 11 foram identificadas como *oxidase alternativa 1*, 22 como *2a* e 14 como *2b*, sendo que a

maior variabilidade foi detetada no gene *2a*. As análises filogenéticas realizadas, não permitiram a identificação de qualquer grupo de sequências associado a qualquer um dos fenótipos em estudo. No entanto, a procura por elementos de regulação realizada nas zonas codificantes em cada um dos genes, detetou três elementos nas sequências do gene *2b*, que estavam associados maioritariamente a linhas embriogénicas. De igual forma, também os resultados do intrão 1 no gene *2b* obtidos a partir do IMeter, apontam este como possível regulador das atividades do gene nessas linhas. Embora limitadas, tais observações apontam para o *2b*, como sendo o gene envolvido na capacidade das células de *Daucus* para desenvolver embriões. No entanto, e dadas as limitações, tal indicação requer uma investigação mais aprofundada.

Palavras-Chave: *Daucus*; Embriogénese somática; Linha celular; Oxidase alternativa; Local polimórfico

Abstract

The somatic embryogenesis (SE) process is the most prominent example of cell reprogramming. *Daucus carota* L. is the first species where totipotency through SE was proven. However, even in an easily inducible plant like *Daucus*, the reprogramming capacity of cells is largely influenced by their genotype. In this view, the identification of markers for 'easy-reprogramming', is expected to help develop functional markers for efficient biotechnological propagation of recalcitrant genotypes. Aiming to explore this issue using alternative oxidase (AOX) as the functional marker gene candidate, 28 *Daucus* accessions were induced to perform SE. Of those, 25 were responsive, producing 139 true-to-type cell lines. SE efficiencies were evaluated throughout a two-step phenotyping method planned for the selection of stable phenotypes. After the first phenotyping step, 41 cell lines with stable SE efficiency, were further analysed and characterized concerning genetic diversity, using the cTBP method. From these, 22 were selected for the second phenotyping step. Eight cell lines were identified at the end as highly stable for extreme SE efficiencies. Four were very efficient or very efficient / efficient, and the other four were non-embryogenic. Those were established as the basic collection used for further poidy assessment and molecular analyses of *AOX*. Polyploidy was detected in three of the cell lines, independently of their embryogenic capacity. However, attending the recent studies reporting polyploidy as an important source of variability for general stress tolerance, it was determined to proceed including them. The three *AOX*s were amplified from the collection and explored in view of the identification of polymorphic positions associated with the detected SE phenotype. In total, 290 sequences were amplified, from which, 47 were identified as polymorphic. From those, 11 were identified as *AOX1*, 22 as *2a* and 14 as *2b*, being *AOX2a* the main source of sequence variability. General phylogenetic analysis did not allow the identification of any group of sequences associated with any SE phenotype. Nevertheless, the search for regulatory elements performed for each gene coding region, detected three elements in the *AOX2b* sequences which were enriched in the embryogenic lines. In the same way, also IMEter results obtained from *AOX2b* intron 1 sequences, reveal these as likely candidates to regulate gene activities. Such observations, point to *AOX2b,* as a gene involved in the capacity of the *Daucus* cell's

to develop embryos. However, given the limited observations, further investigation is required to better substantiate this conclusion.

Key-Words: *Daucus*; Somatic embryogenesis; Cell line; Alternative oxidase; Polymorphic site

Abbreviations, acronyms, symbols and molecular formulas

- % - Percentage;
- $(NH_4)_2SO_4$ - Ammonium sulfate;
- °C - Degrees Celsius;
- µl - Microliter;
- µM - Micromolar;
- 1C - Content of DNA;
- 2,4-D - 2,4-diclorophenoxyacetic acid;
- 2n - Diploid number;
- A - Adenine;
- A1 - HVA1 motif element from barley GCCGAC gene;
- ABA - Abscisic acid;
- ABRE - ABA response element;
- ADP - Adenosine diphosphate;
- AFLP - Amplified fragment length polymorphism;
- AOX - Alternative oxidase*;*
- ATP - Adenosine triphosphate;
- AUS - Australia;
- B_5 - Gamborg basal 5 medium;
- *BBM - BABY BOOM;*
- BPC1 - Basic pentacysteine 1;
- BLAST - Basic local alignment search tool;
- bp - Base pair;
- C - Cytosine;
- $C_{10}H_{12}N_2NaFeO_8$ - Ethylenediaminetetraacetic acid ferric sodium salt;
- $C_{12}H_{17}ClN_4OS.HCl$ - Thiamine hydrochloride;
- $C_6H_{12}O_6$ - Myo-inositol;
- $C_6H_5NO_2$ - Nicotinic acid;
- $C_8H_{11}NO_3.HCl$ - Pyridoxine hydrochloride;
- Ca^{2+} - Calcium ion;
- $CaCl_2.2H_2O$ - Calcium chloride dihydrate;
- cm - Centimeter;
- $CoCl_2 .6H_2O$ - Cobalt (II) chloride hexahydrate;

- cTBP - Combinational tubulin-based polymorphism;
- $CuSO_4 .5H_2O$ - Copper (II) sulfate pentahydrate;
- Cys_I - Cysteine I;
- DAPI - 4',6-diamidino-2-phenylindole;
- DArT - Diversity arrays technology;
- Dc or *D. c.* - *Daucus carota*;
- DEU - Deutschland;
- DNA - Deoxyribonucleic acid;
- DNK - Denmark;
- DSE - Direct SE;
- E - Efficient;
- *Eco*RI - *Escherichia coli* restriction endonuclease enzyme 1;
- Em - Early methionine;
- Em1b - Early methionine 1b;
- EPIC - Exon-primed intron-crossing;
- F1 - F1 hybrid cultivar;
- FRA - France;
- G - G box (found in the section 3 and Appendix 9, do not confuse with guanine);
- G - Guanine (found in the section 1 and Appendices 3, 7 and 9, do not confuse with G box);
- g/l - Gram per liter;
- GA - Gibberellic acid;
- GA5 - GA motif 5 element;
- GBR - Great Britain;
- gDNA - Genomic DNA;
- h - Hour;
- H^+ - Oxidized hydrogen;
- H_2O_2 - Hydrogen peroxide;
- H_3BO_3 - Boric acid;
- HRIGRU - Horticulture Research International - Genetic Resources Unit;
- HSP - Heat shock protein;
- HUN - Hungary;
- I - Inefficient;
- ILP - Intron length polymorphism;

- IME - Intron-mediated enhancement;
- InDels - Insertions and deletions;
- IPK - Institute of Plant Genetics and Crop Plant Research;
- IPTG - Isopropyl β-D-1-thiogalactopyranoside;
- ISE - Indirect SE;
- ISR - Israel;
- ITA - Italy;
- JBUL - Lisbon University Botanical Garden;
- JIM8 - John Innes Monoclonal Antibody 8;
- JKI - Julius Kühn - Institut - Federal Research Centre for Cultivated Plants;
- JPN - Japan;
- kb - kilobase;
- kg/cm^2 - Kilograms per square centimeter;
- KI - Potassium iodide;
- KNO_3 - Potassium nitrate;
- l - Liter;
- L - Long *AOX2a* type;
- LB - Luria Bertani;
- LEA - Late embryogenesis abundant proteins;
- *LEC - LEAFY COTYLEDON;*
- M - Molarity;
- MAFFT - Multiple alignment using fast fourier transform;
- MAR - Morocco;
- Mbp - Mega base pairs;
- mg/ml - Milligrams per milliliter;
- $MgSO_4.7H_2O$ - Magnesium sulfate heptahydrate;
- Mha - Mega hectare;
- ml - Milliliters;
- mm - Milimeters;
- mM - Millimolar;
- $MnSO_4.H_2O$ - Manganese (II) sulfate monohydrate;
- MSA - Mitosis-specific activator;
- mTP - Mitochondrial targeting peptide;
- $Na_2MoO_4.2H_2O$ - Sodium molybdate dihydrate;

- NaCl - Sodium chloride;
- NADP - Nicotinamide adenine dinucleotide phosphate;
- $NaH_2PO_4 .2H_2O$ - Sodium dihydrogen phosphate;
- NCBI - National center for biotechnology information;
- NDH - Non-phosphorylating NAD(P)H dehyidrogenase;
- NE - Non-embryogenic;
- ng/ml - Nanogram per milliliter;
- ng/µl - Nanogram per microliter;
- NGB - Nordic Gene Bank;
- NLD - Netherlands;
- nm - Nanometers;
- nsSNP - Non-synonymous SNP;
- Ø - Diameter;
- O - Oxygen;
- OP - Open pollinated cultivar;
- ORF - Open reading frame;
- PCR - Polymerase chain reaction;
- pg - Picogram;
- PGR - Plant growth regulator;
- pH - Power of hydrogen;
- P_i - Inorganic phosphorus;
- POP - Population;
- PP_i - Inorganic pyrophosphatase;
- PRT - Portugal;
- psi - Pound per square inch;
- PSLUR - Plant Science Laboratories, The University of Reading;
- Q - Ubiquinone;
- RAPD - Random Amplified Polymorphism DNA;
- RFLP - Restriction fragment length polymorphism;
- RNA - Ribonucleic acid;
- S - Short *AOX2a* type;
- SA - Salicylic acid;
- SE - Somatic embryogenesis;
- *SERK - SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE*;

- SH - Sequence harmony;
- SNP - Single nucleotide polymorphism;
- SPP - Single plant progeny;
- sSNP - Synonymous SNP;
- SSR - Simple sequence repeat;
- Subsp. - Subspecies;
- T - Thymine;
- T0 - Inoculation time point;
- T15 - Observation and documentation point after 15 days;
- T30- Observation and documentation point after 30 days;
- T45 - Observation and documentation point after 45 days;
- T60 - Observation and documentation point after 15 days;
- TP - Targeting peptide;
- TRX *h* - Thioredoxin *h*;
- TUR - Turkey;
- UCP - Uncoupling protein;
- UPGMA - Unweighted pair group method with arithmetic average;
- URY - Uruguay;
- UTR - Untranslated region;
- UV - Ultraviolet;
- V - Volt;
- v - Volume;
- VE - Very efficient;
- VI - Very inefficient;
- W - W box element;
- w - Weight;
- X-Gal - 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside;
- $ZnSO_4.7H_2O$ - Zinc sulfate heptahydrate;
- λDNA - Lambda DNA;
- μmol m^{-2} s^{-1} - Micromoles of light per square meter per second.

Index

Index of figures

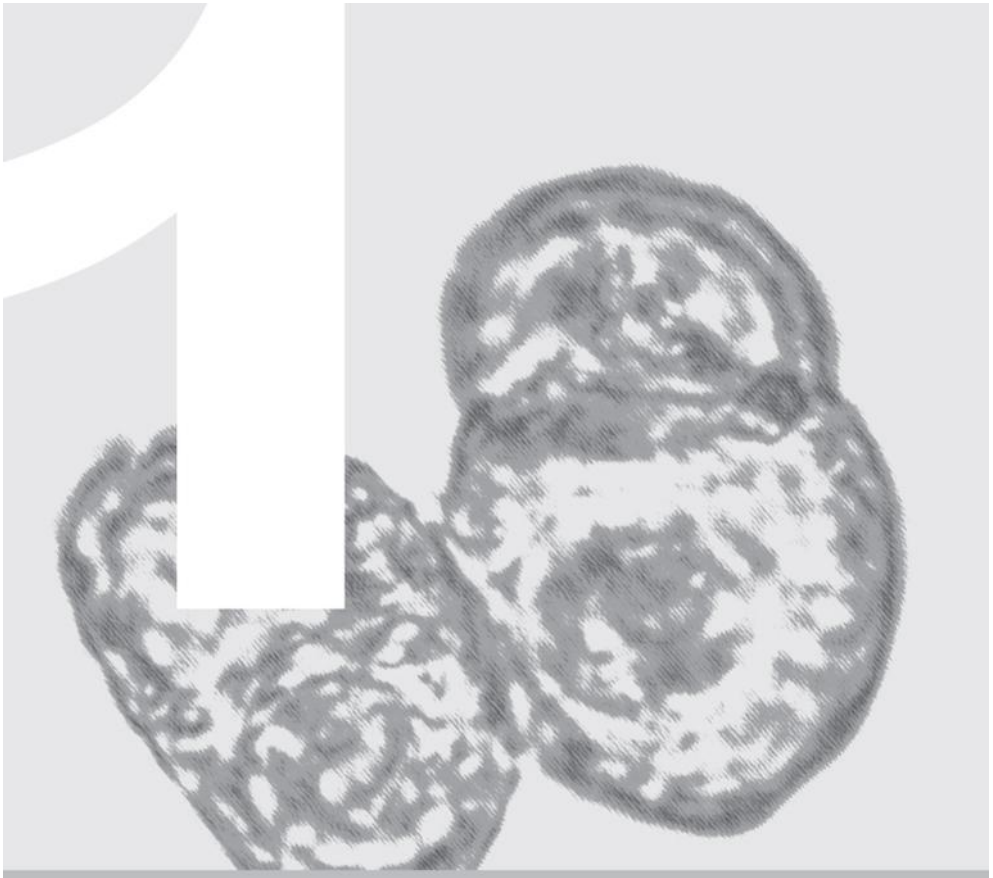
base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 1949 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c*. 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'. 60

Figure 3.9 - Clustering analyses of the 22 *AOX2a* nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 5385 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'. 62

Figure 3.10 - Clustering analyses of the 14 *AOX2b* nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above

line - 2_4 - *D. c.* ‘Senta’; 4_5 - *D. c.* ‘Nantes normu’; 8_2 - *D. c.* ‘Lange rote stumpfe‘; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* ’Rotin’; 27_2 - *D. c.* ‘Nevis F1’. 77

Index of tables



STATE OF THE ART

1 - State of the art

1.1 - *Daucus carota* L.: Classification, botany and agro-biotechnological relevance

The species *Daucus carota* L. (*D. c.*) is a member of the *Daucus* genus, which belongs to the *Apiaceae* (*Umbelliferae*) family. In the last century the genus was described and revised several times (Heywood, 1968 and 1982; Sáenz-Laín, 1981). In addition, several molecular and morphological phylogenetic studies were performed (Lee and Downie, 1999 and 2000; Lee *et al.*, 2001; Spalik and Downie, 2007; for a review see Grzebelus *et al.*, 2011). These works revealed that the genus is not monophyletic as its inclusive branch of *Daucinae* (named *Daucus* senso lato clade) includes four subclades with only two of them traditionally placed in the genus. These are, namely: *Daucus I* and *II* subclades comprising in total, around 21 - 24 species (Spalik and Downie, 2007). The *Daucus I* subclade includes the wild ancestor of the cultivated *Daucus* with all of its subspecies, as well as the several Mediterranean members and some species that were traditionally placed in other genera. The *Daucus II* subclade comprises the remaining members of *Daucus* including its American and Australian representatives (Spalik and Downie, 2007).

All *Daucus* identified so far, are diploid out-crossing species, with a diploid value (2n) of 20 or 22 chromossomes being found in the majority of them (Iovene *et al.*, 2008). Exceptions were found in the species *D. carota*, *D. capillifolius*, *D. sahariensis* and *D. syrticus*, with a 2n value of 18 and for *D. glochidiatus*, the most divergent one, with 44 of 2n value (Imani *et al.*, 2001; Iovene *et al.*, 2008; Iovene *et al.*, 2011). The haploid genome size of *Daucus* had been estimated at 473 Mbp (Arumuganathan and Earle, 1991; Bennett and Leitch, 1995), with a deoxyribonucleic acid (DNA) content value (1C), ranging from 1 to 4,7 pg, depending on the species (Bennett and Smitt, 1976; Grzebelus *et al.*, 2011).

However, due to the referred polyphyly, the current classification system of *Daucus* is untenable (Vivek and Simon, 1999). Grzebelus *et al.* (2011) expressed the need for additional data from molecular and morphological markers in order to develop a more workable classification system. In this view, the works developed by Shim and

Jørgensen (2000), using amplified fragment length polymorphism (AFLP), or Bradeen *et al.* (2002), using several molecular markers and the ones developed by Baranski *et al.* (2012), using simple sequence repeat (SSR) and Iorizzo *et al.* (2013) using single nucleotide polymorphism (SNP) diversity, represent a step forward in the elucidation of *Daucus* domestication and breeding history, which still remains mostly unclear. More recently, Grzebelus *et al.* (2014), developed a diversity arrays technology (DArT) platform for wild and cultivated *Daucus* and used it to investigate genetic diversity and to develop a saturated genetic linkage map using 94 cultivated and 65 wild accessions. As a result, accessions were attributed to three separate groups (wild, Eastern cultivated and Western cultivated) and 27 markers were identified by showing signatures for selection. They showed a directional shift in frequency from the wild to the cultivated, likely reflecting diversifying selection imposed in the course of domestication. This provides a powerful background for further research on the history of *Daucus* domestication.

Daucus members are usually herbaceous biennials, rarely annuals, growing from slender to very stout taproots (IPGRI, 1998). The leaves are pinnatisect, the inflorescence is a compound umbel and the fruit is a schizocarp splitting into two one-seeded mericarps. The fruit is oblong to ovoid, dorsally compressed, with prominent longitudinal projections. Primary ribs are present on each mericarp, situated above vascular bundles, as well as two secondary ribs, situated between the primary ones. Primary ribs are covered with two to four rows of unbranched, semi-erect or spreading hairs. On each secondary rib there is a row of spines, which can be glochidiate or simple at the apex (Grzebelus *et al.*, 2011).

The cultivated *Daucus* is reputed to have its primary origins in Afghanistan and Central Asia, with the Himalayan-Hindu Kush region being the origin of Eastern cultivars and the Anatolian region of Asia Minor being the center of diversity for Western ones (Soufflet-Freslon *et al.*, 2013). *D. carota*, including wild and cultivated forms, has wide phenotypic and molecular variation (Hauser and Bjorn, 2001; Hauser, 2002). Early selection probably focused on biggest, smooth storage and a reduced tendency for early flowering (Stolarczyky and Janick, 2011). The broad variation in cultivated *Daucus* suggests that widespread introgression of wild germplasm has likely occurred into cultivated forms (Simon, 2000; Hauser and Bjorn,

2001; Hauser, 2002). After domestication and dissemination throughout Eurasia, the next known major change on cultivated *Daucus* was the shift in storage root color from yellow and purple to orange in the late 16th and early 17th centuries (Banga, 1957a, 1957b and 1963; Stein and Nothanagel, 1995; Hauser *et al.*, 2004; Umehara *et al.*, 2005; Rong *et al.*, 2010).

The development of *Daucus* cultivars through traditional breeding methods has been a major effort since the 1980's and resulted in significant improvements concerning yield and quality (Simon, 1984; Ammirato, 1986; Peterson and Simon, 1986). The species broad germplasm base has been used regularly in modern *Daucus* breeding. From there, important agronomic traits were introduced into the modern *Daucus* cultivars, such as cytoplasmic male sterility, elevated carotene content and resistance to several diseases and pests (Simon, 2000). According to Punja *et al.* (2007), root shape, length and color, smooth skin, flavor, early maturity, and resistance to various diseases were and continue to be agronomic traits with high priorities in *Daucus* breeding. However, most of these improvements require long-term efforts due to the multigenic control of these traits (Rong *et al.*, 2010). Currently, much of *Daucus* production comprises F_1 hybrids (Luby and Goldman, 2016), produced using a system of cytoplasmic male sterility that makes crossing of inbred lines achievable and economically viable (Allard, 1960; Peterson and Simon, 1986; St. Pierre and Bayer, 1991; Stein and Nothanagel, 1995, Simon, 2000). At present, no transgenic *Daucus* cultivars are available in the market (Punja *et al.*, 2007). In spite of this, the recovery of transgenic *Daucus* plants has already been reported from several laboratories (Takaichi and Oeda, 2000; Peters *et al.*, 2011; Ahn *et al.*, 2012).

Members of the *Daucus* genus are the most widely grown crops of the family *Apiaceae*, cultivated on 1,2 Mha globally (FAO, 2011). *Daucus* cultivars are widely grown worldwide for their edible taproots (Punja *et al.*, 2007). *Daucus* taproots are marketed as fresh whole or baby size, and are used after processing in canned foods, soups or juice, and in frozen products. Nutritionally, *Daucus* are highly rich in β-carotene (provitamin A), as well as vitamin B_1 and C, and are a good source of dietary fiber (Punja *et al*., 2007).

In the biotechnological field, the species gained its honor place with the pioneering works of Steward *et al.* (1958) and Reinert (1958), where *Daucus* was used to demonstrate totipotency of plant cells. Currently, *Daucus* is used as a model species for tissue culture, and extensive work has been conducted in several areas, including somatic embryogenesis (SE), bioreactor scale-up of suspension cultures, protoplast culture, somaclonal variation and pharmacological research (Ammirato, 1986 and 1987; Zimmerman, 1993; Komamine *et al.*, 2005; Shaaltiel *et al.*, 2007; Peters *et al.*, 2011; Rosales-Mendoza and Tello-Olea, 2015).

1.2 - Somatic embryogenesis: Changing fate under stress

The most extreme example of flexibility in plant development is the capacity of several cell types, in addition to the zygote, to initiate embryonic development (Veit, 2006; Sablowski, 2007; Capron *et al.*, 2009). *In vitro* or *in vivo* SE represents this remarkable developmental process, organized in a sequence of stereotypical morphological transformations, enabling non-zygotic plant cells (somatic cells), including haploid cells, to differentiate into somatic embryos and regenerate complete plants, bypassing the fusion of gametes (Rose *et al*., 2010; Nic-Can *et al*., 2015). SE follows a unique development pathway, during which cells have to dedifferentiate, activate cell division, and reprogram their physiology, metabolism and gene expression patterns (Yang and Zhang, 2010; Joshi and Kumar, 2013) and represents the maximum expression of cell totipotency (Gutiérrez-Mora *et al*., 2012). Likewise their zygotic counterpart, somatic embryos pass through four general sequential developmental stages, namely: globular, heart, torpedo and cotyledonary (Figure 1.1) (Yeung, 1995; Dodeman *et al.*, 1997). SE excels beyond other forms of regeneration such as organogenesis in that within a single step it produces a vascular system, functional meristem and a root/shoot axis (Bassuner *et al.*, 2007). Somatic embryos can be differentiated either directly, from the explants without an intervening *callus* phase, or indirectly, after a *callus* phase, referred to as direct SE (DSE) and indirect SE (ISE), respectively (Sharp *et al*., 1980; Quiroz-Figueiroa *et al*., 2006). Uni- or multicellular pathways have been identified as the origin of somatic embryos in both, ISE or DSE (Quiroz-Figueiroa *et al*., 2006).

SE has been a very valuable tool for achieving a wide range of purposes, from the basic biochemical, physiological and morphological studies, to the development of technologies with a high degree of practical applications. The investigation of initial events of plant embryogenesis (Zimmerman, 1993; Yu *et al.*, 2006), the mass production of plants (Bonga *et al.*, 2010; Ji *et al.*, 2011; Nic-Can *et al*., 2015), synthetic seeds (Kumar, 2000; Reddy *et al.*, 2012), polyploids production (Lee *et al.*, 2009a; Sun *et al.*, 2011; Delporte *et al.*, 2012; Konieczny *et al.*, 2012), protoplast source (Jiang *et al.*, 2013), gene transfer for crop improvement (Santos *et al.*, 2002; Kamle *et al.*, 2011; Delporte *et al.*, 2013), secondary metabolite production (Vanisree *et al.*, 2004; Hussain *et al.*, 2012; Rahmawati and Esyanti, 2014) and toxicity

screening tests (Chen and Punja, 2002; Pérez-Clemente and Gómez-Cadenas, 2012), are a few technological examples where SE has been widely applied.

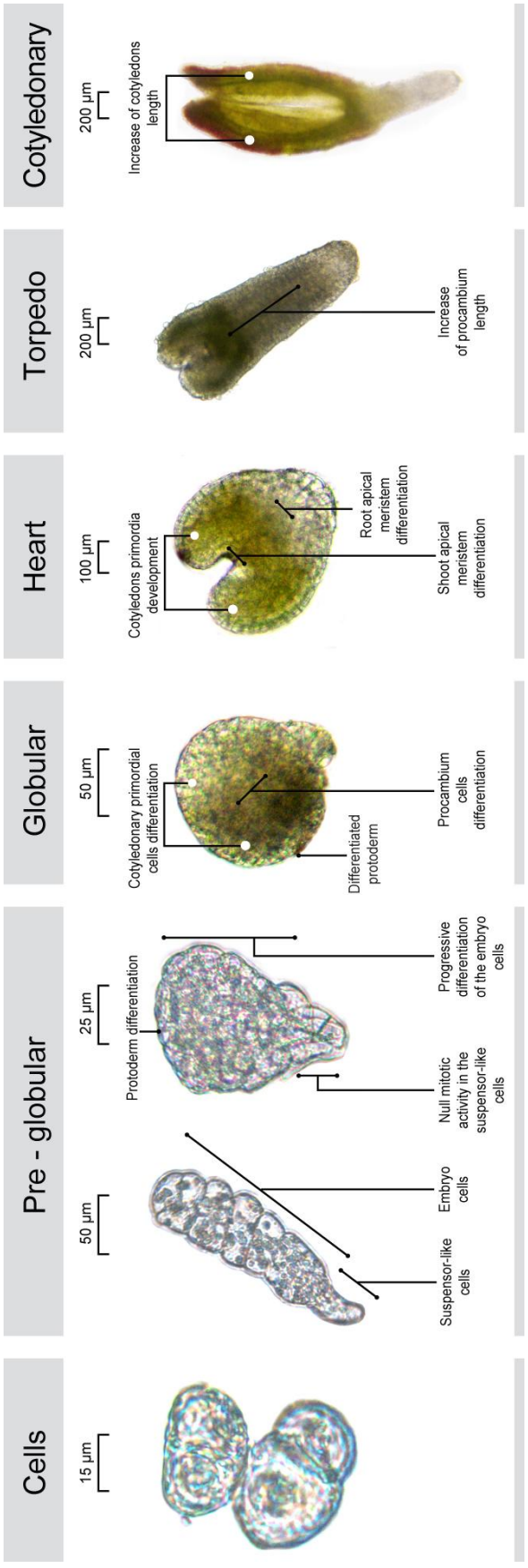


Figure 1.1 - Developmental stages of somatic embryos. Adapted from Frederico *et al.* (2009a).

Since the prediction of cell totipotency by Haberlandt in the early 1900's and the pioneering works from Reinert (1958) and Steward *et al*. (1958), much progress in SE understanding has been made, using the *Daucus* model system. Since these first studies, the number of higher plant species from which somatic embryos could be obtained and regenerated, has continuously increased. The phenomenon has been documented in a large number of gymnosperm and angiosperm species (Raemakers *et al*., 1995; George *et al.*, 2008). Some species, however, are more recalcitrant than others regarding both, the initiation of embryogenic cultures and the regeneration of plants (Rao, 1996; Li *et al.*, 2006; Nic-Can *et al*., 2015). The successful induction of somatic embryos and subsequent recovery of viable plants is not routine or efficient for the majority of species (Merkle *et al*., 1995) and the mechanisms involved in the induction and establisment of SE remains mostly unknown (Jiménez, 2001).

Determining specific physical (species, explants origin and environmental conditions) and chemical factors, such as culture media, type/concentration of plant growth regulators (PGRs) and nitrogen/carbon source, that switch on the development of the embryogenic pathway remains a key step in embryogenic induction (Elmeer, 2013). Early research on SE mostly focused on PGRs (Jiménez, 2001; Raghavan, 2004), and a repertoire of strategies has been developed to regenerate many species via SE using PGRs as inducers (Yang and Zhang, 2010). Finding the right conditions to induce SE in different species and cultivars has been mostly based on trial and error experiments (Jacobsen, 1991; Henry *et al*., 1994), by analyzing the effect of different culture conditions and media and modifying especially the type and levels of PGRs (Jiménez, 2001). Auxins and cytokinins have been considered to be the most important PGRs in relation to cell division and differentiation, as well as in the induction of SE (Fehér *et al*., 2003). The auxin analog herbicide 2,4-diclorophenoxyacetic acid (2,4-D), has been widely used, especially in the latter process (Dudits *et al*., 1991; Yeung, 1995; Fehér *et al.*, 2002). A large amount of *in vitro* SE systems rely on the use of exogenous 2,4-D as an inducer (Nomura and Komamine, 1995; Pedrosa and Vasil, 1996; Meneses *et al.*, 2005; Sharma *et al.*, 2005; Lee *et al.*, 2009b; Sharifi *et al*., 2012 and references therein). However, embryo development in somatic tissues has been reported in the absence of PGRs, as well as in the presence of other PGRs, such as cytokinins (Eudes *et al.*, 2003; Gaj, 2004; Jia *et al.*, 2008), gibberellic acid (GA) (Swain *et al.*, 1997; Hay *et al.*, 2002;

Wang *et al.*, 2004; Nasim *et al.*, 2010) or abscisic acid (ABA) (Nishiwaki *et al.*, 2000; Ikeda *et al.*, 2004 and 2006; Kikuchi *et al.*, 2006; Rai *et al.*, 2011; Jin *et al.*, 2014). SE can also be promoted by non-hormonal inducers, such as high sucrose concentration, or osmotic stress (Kamada *et al*., 1993; Ikeda *et al*., 2004), heavy metal ions (Kiyosue *et al*., 1990; Pasternak *et al*., 2002), high temperature (Kamada *et al*., 1989; Kikuchi *et al*., 2006; Fu *et al.*, 2008; Aslam *et al.*, 2011) and light (Torné *et al.*, 2001; Germanà *et al*., 2005). It has been proposed that PGRs and stress play a central role in mediating the signal transduction cascade leading to the reprogramming of gene expression, followed by a series of cell divisions that induce either unorganized *callus* growth, or polarized growth, leading to SE (Dudits *et al*., 1991). These findings have prompted the question of whether SE is a stress response of plants to survive extreme *in vitro* environmental conditions (Fehér *et al.*, 2002, Ikeda *et al.*, 2004; Karami and Saidi, 2010; Zavattieri *et al.*, 2010; De-la-Peña *et al.*, 2015). In fact, several reports support this point of view. Jin and co-workers (2014) used sodium chloride (NaCl) and ABA stress treatments in *Gossypium hirsutum* to regulate the balance between proliferation and differentiation that determines SE development. Potters *et al.* (2007), Potters *et al.* (2009) and lately Grafi *et al.* (2011), proposed that a common response of plant cells to sub-lethal stress is cellular dedifferentiation, whereby cells first acquire a stem-like state before assuming a new fate, which represents the first steps of embryogenic commitment and lately the acquisition of an embryogenic state leading to the SE pathway development (Verdeil *et al.*, 2007; Zavattieri *et al.*, 2010). The cell state shift from somatic into a embryogenic state is accompanied by the synthesis of ribonucleic acid (RNA) and DNA, a change in pH, an increase in the rate of oxygen uptake, elevated enzyme activity (mainly kinase), migration of nuclei towards the cell wall, changes in the cytoskeleton, active conversion of adenosine triphosphate (ATP) to adenosine diphosphate (ADP), inactivation of cytosolic factors and maturation promotion factor (Karami *et al.*, 2009; Kurczyńska *et al.*, 2012). These SE committed cells are usually isodiametric, rich in cytoplasm and starch, with callose depositions, being separated from the rest of the cells after severed plasmodesmata process (Suprasanna and Bapat, 2005).

Most of the success achieved so far in understanding the mechanisms that govern the efficient regeneration of plants through SE has been accomplished with model

plant species, such as *Daucus* (Vries *et al*., 1988; Lin *et al.*, 1996; Imani *et al*., 2002; Fujimura, 2014), *Medicago* (Fowler *et al*., 1998; Fehér *et al*., 2002) and *Arabidopsis* (Ikeda-Iwai *et al*., 2003; Raghavan, 2006). However, the transfer of these new technologies to major crop species has been slow and difficult (Vasil, 1987; Raghavan, 2004). In this way, independently of the nature of the external stimulus, the establishment of SE necessarily involves profound changes at the molecular level, such as the coordinated expression of different sets of genes that drive the switch from vegetative grown to embryogenic development (Rose and Nolan, 2006). Thus, the identification of the genes that trigger key phases of SE, i. e. cell dedifferentiation, cell cycle reentry and establishment of a new embryogenic fate, has been highly desirable (Thomas and Jiménez, 2005; Fehér, 2008).

Since there are no universal cytological markers by which somatic embryogenic cells can be distinguished from non-embryogenic ones, the interest of biologists has long turned to invisible molecular markers, seeking to identify genes regulating SE-triggered changes (Elhiti *et al.*, 2013). The identification and characterization of gene markers for SE offer the possibility of determining the embryogenic potential of somatic cells before any morphological changes appear and to provide information on molecular regulation of early SE (Mahdavi-Darvari *et al.*, 2015). With the advent of genomics, numerous studies have been conducted to identify genes responsible for the various stages of SE (Chugh and Khurana, 2002; Suprasanna and Bapat, 2005; Chugh and Eudes, 2007), using model plant species and others. Since the beginning the *Daucus* system has been and continues to be widely used (Bayliss, 1976; Chibbar *et al*., 1988; Coutos-Thevenot *et al.*, 1990; De Jong *et al.*, 1993; Dyachok *et al.*, 2000; Imani *et al*., 2001; Li and Kurata, 2005; Imani *et al*., 2006; Frederico *et al*., 2009a; Peters *et al*., 2011).

The idea that PGRs may be perceived as a stress condition resulted from work on *Daucus* cells, where two heat shock protein (HSP) genes were found to be auxin-responsive during somatic embryo development (Coca *et al*., 1994; Kitamiya *et al*., 2000). Cell tracking has been successfully applied to evaluate and mark the fate of embryogenic cells using the John Innes Monoclonal Antibody 8 (JIM8) reactive cell wall epitope (Pennell *et al.*, 1992; Pennell *et al.*, 1995, McCabe *et al.*, 1997), and to elucidate the signaling pathways by which plant cells remodel their gene expression

program (Souter and Linsey, 2000; Jiménez and Thomas, 2006). The identification of hormone-inducible genes has also yielded clues how regulation of gene expression is controlled during embryogenic development. The characterization of signaling component genes, such as *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*) (Schmidt *et al*., 1997), has generated great interest in the switching of several signaling cascades during SE, uncovering transcription factors such as *BABY BOOM* (*BBM)*, *LEAFY COTYLEDON* (*LEC1* and *2*) as potential regulators of SE development (Yang and Zhang, 2010).

Although much is already known, it remains unclear what underlies the differences in SE-mediated regeneration efficiency observed amongst different genotypes within the same species and the even more drastic differences between species (Ochatt *et al*., 2010). The observation that different genetic mutations induce similar embryogenic phenotypes in postembryonic plants, reflects the complexity of SE and the possible existence of overlapping pathways triggering the SE developmental process, under the local tissue/cellular conditions and appropriated hormonal balance (Thomas and Jiménez, 2005). This variability in competence for *in vitro* regeneration via SE has handicapped and delayed the exploitation of biotechnology approaches for breeding in many species and still is the main research subject in several laboratories (Ochatt *et al*., 2010). When different genotypes are used, differential responses are observed even in *Daucus*, one of the most extensively studied species concerning SE (Wilde *et al.*, 1988; Imani *et al*., 2001; Frederico *et al*., 2009a). However, the molecular basis underlying these differential responses remain unknown (Karami *et al*., 2009).

Efforts to overcome this situation were made by comparing embryogenic and non-embryogenic cells at multiple levels, such as at the morphological, genomic and proteomic (Sharifi *et al*., 2012; Nic-Can *et al*., 2015). At the morphological level embryogenic *calli* were described as presenting nodular features, with a friable and smooth surface (Yang and Zhang, 2010), being highly variable in color depending on the species, ranging from brown in *Coffea arabica* (Quiroz-Figueroa *et al*., 2006) to translucent in *Crocus sativus* (Sharifi *et al.*, 2012) or yellow/light yellow in *Gossypium hirsutum* (Han *et al*., 2009) and *Daucus* (Frederico *et al.*, 2009a). In contrast, according to the same authors, non-embryogenic *calli* were always described as

presenting a rough surface and being usually hard or spongy, with dark colors (green to brown).

On the other hand, the determination of *calli* embryogenic efficiency for regeneration was mostly performed by methodologies based on embryo counting amongst genotypes in order to select the most efficient for further evaluation (Lin *et al.*, 1996; Han *et al.*, 2011). This approach is usually highly time consuming and tedious, may thus be the reason why SE efficiency evaluations only consider a few genotypes and even a lower number of *calli* cell lines, without considering the differences within accessions (Han *et al.*, 2011). This limited the achievement of a global overview and broad conclusions in relation to the differential SE response (Sujatha, 2011), when genomic and proteomic studies are performed to identify the factors leading to embryogenic progression or repression (Zeng *et al*., 2007; Yang and Zhang, 2010). According to Ochatt *et al.* (2010) and Elmeer (2013), the identification and characterization of the embryogenic capacity of a specific species/genotype/accession, will increase the accuracy of the acquired conclusions and will lead to a better understanding of the differential SE efficiency responses. On the other hand, the use of germplasm from model species, with well defined SE protocols and an 'easy-to induce' capacity, such as in the case of *Daucus*, is expected to increase the possibility of success in the search for factors leading to an improved embryogenic efficiency for plant regeneration, especially in the recalcitrant ones.

1.3 - Alternative oxidase: Recycling pump and much more

Plant alternative oxidase (AOX) is a multigene family encoded in the nucleus by two discrete subfamilies (*AOX1* and *AOX2*), and was firstly discovered in angiosperms upon examination of the phenomenon of cyanide-resistant respiration (Bendall and Bonner, 1971). Family members have been identified in all higher plants investigated to date and also in some algae, fungi (Scheckhuber *et al.*, 2011), eubacteria and protists (Whelan *et al.*, 1996; Baurain *et al.*, 2003; Stenmark and Nordlund 2003; Venter *et al.*, 2004; McDonald and Vanlerberghe, 2006). The presence of AOX was also revealed in animal kingdom phyla, including mollusca, nematoda and chordate (McDonald. and Vanlerberghe, 2004 and 2006; McDonald, 2008; McDonald *et al.*, 2009).

AOX is localized to the inner mitochondrial membrane (Figure 1.2) and is a member of the diiron carboxylate group of proteins, characterized by an active site that includes two iron atoms coordinated by several highly conserved glutamate and histidine residues (Berthold and Stenmark, 2003). AOX is of research interest for studying the phenomenon of retrograde signaling between the mitochondrion and the nucleus and due to its role in the acclimation of plants to a variety of environmental stressors (Gray *et al.*, 2004; McDonald, 2008; Giraud *et al.*, 2009). The role of mitochondria as a physical platform for biochemical networks, signal perception and signal transduction, was proven to be crucial for the maintenance of homeostasis in plants (Raghavendra and Padmasree, 2003; Fernie *et al.*, 2004; Amirsadeghi *et al.*, 2007; Noctor *et al.*, 2007; Rhoads and Subbaiah, 2007; Sweetlove *et al.*, 2007). As part of a global mitochondrial response, Rasmusson *et al.* (2009), Vanlerberghe *et al.* (2009) and recently Chocobar-Ponce *et al.* (2014), suggested the enrollment of AOX in counteracting deleterious short-term metabolic fluctuations, especially under stress conditions by acting as a stress-signaling pathway from the mitochondrion that controls cellular responses to adverse conditions.

AOX catalyzes the oxidation of ubiquinol and reduction of oxygen to water. Hence, the electron transport chain is branched, such that electrons in the ubiquinone pool are passed to oxygen (O_2) via either the cytochrome pathway (using complex III, cytochrome c and cytochrome oxidase) or AOX. The AOX branch of respiration is

non-energy conserving and while its physiological role is still a matter of debate, a developing idea from plant studies is that it may act to dampen the rate of electron transport chain-generated reactive oxygen species (Finnegan *et al*., 2004).

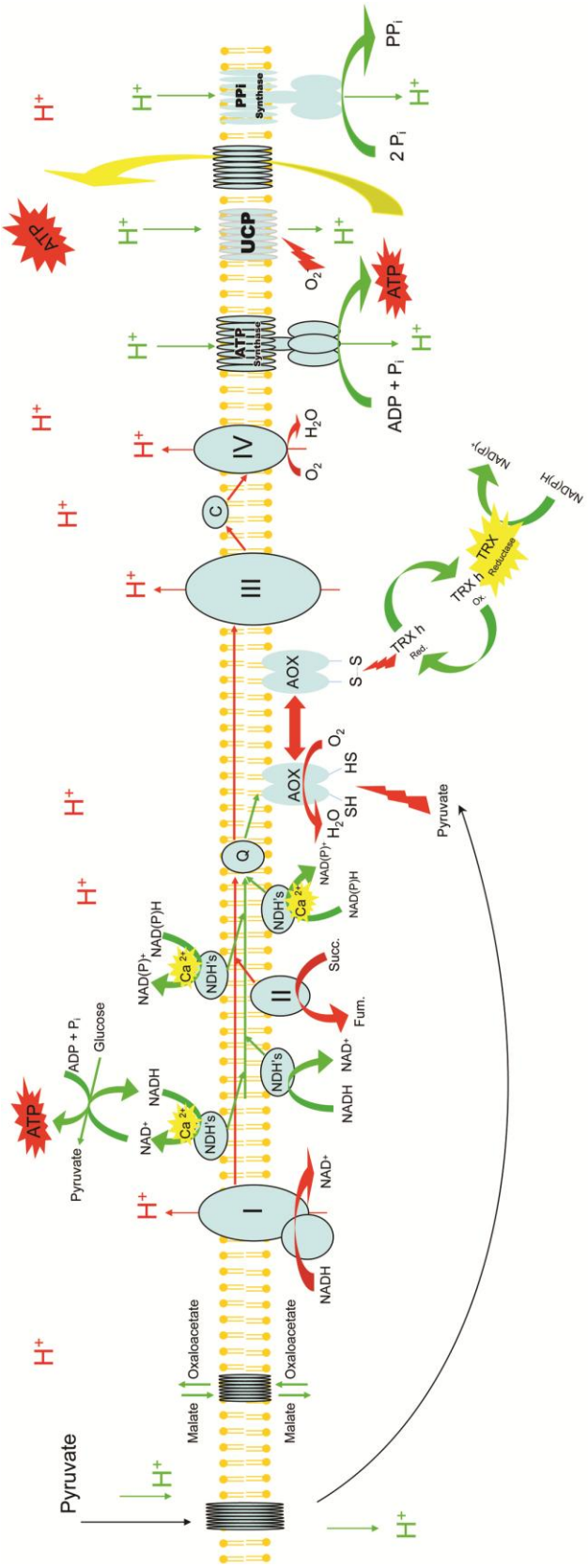


Figure 1.2 - Organization of plant mitochondrial electron transport chain. Electron flux to and from Q is represented according to Fernie *et al.* (2004) and Plaxton *et al.* (2006). **I** - Nicotinamide adenine dinucleotide phosphate (NADP) dehydrogenase; **II** - Succinate dehydrogenase; **III** and **IV** - Cytochrome pathway; **V** - ATP synthase; **c** - Cytochrome c; **ADP** - Adenosine diphosphate; **AOX** - Alternative oxidase; **ATP** - Adenosine triphosphate; **Ca^{2+}** - Calcium ion; **Fum.** - Fumarate; **H^+** - Oxidized hydrogen; **H_20** - Water; **NDH's** - Non-phosphorylating NAD(P)H dehyidrogenases; **P_i** - Inorganic phosphorus; **PP_i** - Inorganic pyrophosphatase; **PPI** - Inorganic pyrophosphatase synthase; **Q** - Ubiquinone; **S** and **SH** - Reduced and oxidized disulfide bonds; **Succ.** - Succinate; **TRX *h*** - Thioredoxin *h*; **UCP** - Uncoupling protein.

The function of AOX in non-thermogenic tissues remains puzzling, possibly due to the wide range of conditions that result in its induction. For instance, AOX protein levels can be induced by several treatments such as chilling (Vanlerberghe and McIntosh, 1992; Purvis and Shewfelt, 1993; Sugie *et al.*, 2006; Mizuno *et al.*, 2007; Feng *et al.*, 2008; Wang *et al.*, 2012), heat (Murakami and Toriyama, 2008), drought (Bartoli *et al.*, 2005; Wang and Vanlerberghe, 2013), osmotic stress (Ederli *et al.*, 2006; Costa *et al.*, 2007) and pathogen attack (Simons *et al.*,1999; Maxwell *et al.*, 2002; Ordog *et al.*, 2002), in addition to treatment with salicylic acid (SA) (Djajanegara *et al.*, 2002; Maxwell *et al.*, 2002), hydrogen peroxide (H_2O_2) (Vanlerberghe and McIntosh, 1996) or with inhibitors of the respiratory chain (Lambers, 1997; Vanlerberghe *et al.*, 1994; Yip and Vanlerberghe, 2001; Vanlerberghe *et al.*, 2002; Mariano *et al.*, 2008; Naydenov *et al.*, 2008). In general, any condition that inhibits or decreases the activity of the main respiratory chain induces the alternative pathway (Djajanegara *et al.*, 2002; Moore *et al*., 2002; Gray *et al.*, 2004; Polidoros *et al.*, 2005; Feng *et al.*, 2013). *AOX1* is most widely known for its induction by stress stimuli in many tissues and is present in both, monocot and eudicot species. *AOX2*, on the other hand, is usually constitutive or developmentally expressed in eudicot species, but is absent from the genomes of all monocot species examined to date (Considine *et al.*, 2002). The gene number of *AOX* in angiosperms is very diverse, ranging from 1 to 6 genes and is comprised of variable combinations of different *AOX* subfamilies and types among species (for review see Costa *et al.*, 2014a)

In the past years, several reports supported the idea that AOX may act during oxidative stress attenuation (Fiorani *et al.*, 2005; Fung *et al.*, 2006; Feng *et al.*, 2008; Giraud *et al.*, 2008). However, the critical importance of the enzyme during acclimation upon stress of plant cells is not fully understood and is still an issue of intensive research and discussion. Clifton *et al*. (2005 and 2006), as well as Arnholdt-Schmitt *et al*. (2006), pointed to the importance of this pathway as an early sensoring system for cell programming. AOX is the critical component in the alternative pathway that transfers electrons from reduced ubiquinone directly to oxygen. The enzyme is related to all types of abiotic and biotic stress and is known to be involved in growth responses and development (Sieger *et al.*, 2005; Umbach *et al.*, 2005; Ho *et al.*, 2007; Sugie *et al.*, 2007; Giraud *et al.*, 2008). A role of AOX was suggested for

the ability of plant cells to change easily their fate upon stress (Frederico *et al.*, 2009a; Zavattieri *et al.*, 2010; Afuape *et al.*, 2013). AOX is increasingly the focus of research on stress acclimation and adaptation and seems to play a key role in regulating the process of cell reprogramming by improving metabolic transitions related with the cellular redox state and the flexible carbon balance (Arnholdt-Schmitt *et al.*, 2006; Rasmusson *et al.*, 2009).

Recently, *AOX* became of central interest as a gene candidate for functional marker development, related to breeding programs focused on improving plant stress responses (Arnholdt-Schmitt *et al.*, 2006; Clifton *et al.*, 2006; Arnholdt-Schmitt, 2009; Polidoros *et al.*, 2009). Several reports presented data concerning polymorphic sites within *AOX* genes, which could have some relevance on gene regulation related to cell reprogramming upon stress. Abe *et al.* (2002) reported the existence of a SNP between alleles of *Oryza sativa AOX1a*, that was tightly linked to the presence of the quantitative trait loci (QTL) for low temperature tolerance. Cardoso *et al.* (2009), Costa *et al.* (2009 and 2014b), Ferreira *et al.* (2009), Frederico *et al.* (2009b) and Santos Macedo *et al.* (2009), reported the existence of several polymorphic sites within *AOX* with potential for gene regulation on several gene regions, including exons, introns and untranslated regions (UTRs). These observations reinforced the strength of *AOX* as a potential marker candidate, because the existence of polymorphisms within gene regions with fully characterized function, is a prerequisite for functional marker development (Andersen and Lübberstedt, 2003; Arnholdt-Schmitt, 2004; Arnholdt-Schmitt, 2005; Arnholdt-Schmitt *et al.*, 2006). However, the understanding the functional relevance of these polymorphic sites and their application on plant breeding programs, requires a long effort and the development of innovative approaches to study the complex relations and functions resulting from genes variability (Agarwal *et al.*, 2008; Poczai *et al.*, 2013).

1.4 - Exploring *Daucus AOX* polymorphisms in the view of functional marker development

In recent years, many promising new alternative molecular marker techniques have been developed in plant genetics and breeding, largely due to rapid growth in genomic research, initiating a trend away from random DNA markers towards gene-targeted functional markers (Andersen and Lübberstedt, 2003; Gupta and Rustgi, 2004; Poczai *et al.*, 2013).

Usually, DNA markers are defined as a fragment of DNA revealing mutations/variations, which can be used to detect polymorphism between different genotypes or alleles of a gene for a particular sequence of DNA in a population or gene pool (Agarwal *et al.*, 2008). Simplifying, a DNA marker is a small region of DNA sequence showing polymorphism (base deletion, insertion or substitution) between different individuals (Jiang, 2013). DNA markers have been developed into various systems based on different polymorphism-detecting techniques or methods, such as, southern blotting-nuclear acid hybridization, polymerase chain reaction (PCR) and DNA sequencing (Collard *et al.*, 2005). Restriction fragment length polymorphism (RFLP), AFLP, Random Amplified Polymorphism DNA (RAPD), SSR and SNP, are examples of other widely used methods (Jiang, 2013).
According to Jiang (2013), and depending on the application and species involved, an ideal DNA marker for efficient use in breeding should meet the following criteria:

- High level of polymorphism;
- Even distribution across the whole genome (not clustered in certain regions);
- Co-dominance in expression (so that heterozygotes can be distinguished from homozygotes);
- Clear distinct allelic features (so that the different alleles can be easily identified);
- Single copy and no pleiotropic effect;
- Low cost to use (or cost-efficient marker development and genotyping);
- Easy assay/detection and automation;
- High availability (un-restricted use) and suitability to be duplicated/multiplexed (so that the data can be accumulated and shared between laboratories);
- Genome-specific in nature (especially with polyploids);

- No detrimental effect on phenotype.

DNA markers developed based on SNPs are the most widely used in plants, followed by insertions and deletions (InDels) (Păcurar *et al.*, 2012; Yamaki *et al.*, 2013). An SNP is a single nucleotide base difference between two DNA sequences or individuals. SNPs can be categorized according to nucleotide substitutions either as transitions (cytosine - C / thymine - T or guanine - G / Adenine - A) or transversions (C / G, A / T, C / A or T / G). In practice, single base variants are considered to be SNPs as are single base InDels in the genome. SNPs provide the ultimate/simplest form of molecular markers as a single nucleotide base is the smallest unit of inheritance, and thus they can provide maximum markers. In plants, SNP frequencies are typically in a range of one SNP every 100 - 300 base pairs (bp) (Edwards *et al.*, 2007; Wu *et al.*, 2010). SNPs may be present within coding, non-coding or intergenic regions between genes at different frequencies in different chromosomes. SNPs are co-dominant markers, often linked to genes and represent the simplest/ultimate form for polymorphism, and thus they have become very attractive as potential genetic markers in genetic studies and breeding programs. Therefore, it can be expected that SNPs will be increasingly used for various purposes, particularly as whole DNA sequences become available for more and more species. Nevertheless, high costs for start-up or marker development, high-quality of the required DNA and high technical/equipment demands, may limit the application of SNPs in some laboratories and practical breeding programs (Jiang, 2013).

On the other hand, functional markers are functionally characterized DNA markers derived from sequence motifs affecting phenotypic variation (Poczai *et al.*, 2013). Functional markers, owing to complete linkage with trait *locus* alleles, are superior to random DNA markers such as RFLPs, SSRs and AFLPs, but require sequences of functionally characterized genes, from which polymorphic functional motifs affecting plant phenotype can be identified. The starting point of any functional marker development is the existence of a gene sequence with an assigned function.

Concerning *AOX*, the development of functional markers using the gene as a candidate, was firstly reported by Arnholdt-Schmitt *et al.* (2006), supported by the

idea based on the metabolic role of alternative respiration under stress, the link between AOX activity and differential growth, and the polymorphisms recently observed in *AOX* genes and reviewed recently by Cardoso and Arnholdt-Schmitt (2013), in the view of a step by step functional marker strategic development for selected traits.

In this view, information concerning the *AOX* genomic structure in plants had been extensively reported, and had been recently reviewed by Cardoso *et al.* (2015), based on data collected from public databases, reporting that the most common gene structure of *AOX* comprises four exons interrupted by three introns. Genes sharing this structure usually present exon size conservation for the last three exons (129, 489 and 60 bp, respectively) (Campos *et al.,* 2009). Size variability of *AOX* encoded by genes with four exon structure is mainly associated with exon 1. Nevertheless, exon size variability can also be observed in the last three exons of *AOX* members. Events of loss, or gain of introns, which have taken place during evolution, are responsible for modifications in the structure of *AOX* and consequently for the changes in exon size (Polidoros *et al.,* 2009). Other known examples are the loss of intron 2 in *AOX1d* and *AOX1b* of *A. thaliana* and intron 3 in *AOX1a* of *Solanum tuberosum*, which resulted into genes with a different structure composed by a longer exon (Considine *et al.,* 2002; Polidoros *et al.,* 2009).

At the protein level, AOX presents highly conserved sites across organisms from diverse kingdoms in both AOX1 and AOX2 subfamily members. Those sites are involved in the coordination of the diiron centre of the enzyme (Siedow *et al.*, 1995; McDonald, 2008), in AOX activity (Moore and Albury, 2008), in anchoring the enzyme to the inner mitochondrial membrane (Crichton *et al.*, 2005 and 2010) and in the catalytic cycle in respect to its interactions with oxygen (Moore *et al.*, 2008). Conserved sites are also located in a hydrophobic region thought to play a role in ubiquinol binding (Albury *et al.*, 2009). Holtzapffel *et al.* (2003) were the first to report variations in the protein functional sites across species (including angiosperms and gymnosperms). The conserved cysteine I (Cys_I), in the N-terminal region, of the protein appeared as serine I in some plant species, in which angiosperms and gymnosperms are included. This substitution consequently changes the enzyme

regulation, which instead of being regulated by pyruvate, is regulated by succinate (Holtzapffel *et al.*, 2003; Grant *et al.*, 2009).

For *Daucus* species, *AOX* was firstly reported by Costa *et al.* (2009 and 2014b). The authors reported the identification of three gene members from the *Daucus* genome, one belonging to *AOX1* and two to the *AOX2*. *Daucus AOX*s characterization had been published by Campos *et al.* (2009) and Cardoso *et al.* (2009), as well as, by Frederico *et al.* (2009a), using diverse biological systems and plant material.

Campos *et al.* (2009) reported the characterization of *Daucus* AOX amino acid sequences and presented data supporting its expression, in all tested tissues and in an *in vitro* primary culture system, attesting its differential responsiveness during *Daucus* biological development.

In the same way, also Frederico *et al.* (2009a) showed that *AOX1* and *AOX2a* are differentially expressed during SE expression in *Daucus*. Additionally, the authors attested its relevance during SE development, by blocking embryo development through the use of the AOX inhibitor salicylhydroxamic acid (SHAM), during the SE expression phase.

On the other hand, the assessment of variability within *AOX* genes regions on *Daucus* was firstly reported by Cardoso *et al.* (2009), with the characterization of *AOX2a* intron 3. An intron length polymorphism (ILP) was identified, leading to the existence of an allele 286 bp longer. The same authors also reported that the presence of SNPs and InDels, was higher in introns when compared to exonic sequences in the same gene, being intron 3 the most affected. Recently, Macko-Podgorni *et al.* (2013) discovered that the ILP previously reported in *Daucus AOX2a* was due to the insertion of a *Stowaway* transposable element. ILP occurrence was also reported in *AOX2b*, although in this case in intron 1, contrarily to what was reported in *AOX2a* (Cardoso *et al.*, 2011). In line with this observation, it is often referred that introns most proximal to the 5' end of a gene are the ones that exert a more pronounced effect on expression (Breviario *et al.*, 2008; Rose, 2008), which increased the importance of this observation in *Daucus AOX2b*. Nevertheless, introns certainly impose a huge energetic burden to the cell, considering that the

density of introns (i.e., the genic regions consuming large amounts of energy for nothing in terms of protein synthesis) is greater than that of exons in genomes. The reasons why introns propagated in some eukaryotic genomes regardless of this energetic disadvantage, are yet to be elucidated. According to Lynch (2002), introns are just selfish DNAs that invade protein-coding genes in eukaryotic genomes, and the deleterious introns can be sustained due to severe population bottlenecks. Many studies have discussed selective advantages that introns bring to the cell in eukaryotes, contributing to overcoming the energetic disadvantage (Gilbert, 1978; Chorev and Carmel, 2012; Jo and Choi, 2015). However, the results derived from such studies are so far controversial (Gilbert, 1985; for review see Gorlova *et al.*, 2014). Despite the different conclusions referred in literature, several reports showed that introns are involved in gene regulation activities, acting for example as enhancers (Mascarenhas *et al.*, 1990; Moabbi *et al.*, 2012), by harboring important regulatory elements (Schauer *et al.*, 2009; Parra *et al.*, 2011). The use of bionformatic tools and the available databases for the search of those putative intronic elements, had already produced important advances in the discovery of important intronic elements and motifs from several species (Morita *et al.*, 2012; Gallegos and Rose, 2015; Pu *et al.*, 2015). These findings opened a new opportunity for the exploitation of the *AOX* intron variability in the view of marker development at the systems and species level, which at the present remains fragmented and unclear. The application of these methodologies on *AOX* is expected to produce additional information concerning the gene regulatory activities related with the existence of polymorphic sites.

In the same way, also the existence of pre-microRNAs has been predicted in *Daucus* intronic regions of *AOX*, such as in *AOX2a* (Cardoso *et al.*, 2009) and *AOX2b* (Cardoso *et al.*, 2011), reinforcing the need of a full assessment of all gene regions.

Recently, Campos *et al.* (2016) presented the characterization of the complete structure of *AOX1* in *Daucus* and Nogales *et al.* (2016), presented new data concerning its variability by studying its allelic variation in different materials, including commercial cultivars, inbred lines, subspecies and wild relatives. Sequence comparison revealed the existence of a high number of SNPs, as well as InDels, especially in exon 1 and intron 1. Intron 1 showed to be the most polymorphic region

and harbored an insertion event of 400 bp, which had highly divergent sequences depending on the *Daucus* genotype. The insertion was located in a region of single tandem repeats that was also polymorphic between genotypes.

Nevertheless, despite the several hits achieved from expression and biochemical data, till the present, no information concerning *Daucus* agricultural traits was referred in relation to the reported *AOX* polymorphic positions. This observation reinforces the necessity for a deeper functional characterization of polymorphic positions. The validation of protocols for the study of *AOX* polymorphic positions in association with a comprehensive and well oriented breeding program, combined with large scale polymorphism search in order to perform QTL association studies, as referred by Nogales *et al.* (2015), will certainly increase the accuracy of the acquired results.

1.5 - Goals

SE is the most prominent example of cell reprogramming. *D. carota* has been the first species where totipotency through SE had been proven. However, even in an easily inducible plant like *Daucus* the reprogramming capacity of cells is largely influenced by their genotype. In this view, the identification of markers for 'easy-reprogramming' is expected to help developing functional markers for efficient biotechnological propagation of recalcitrant genotypes. Considering the stated, this study specifically intends to:

- Develop a new approach to evaluate SE induction and expression efficiency using a large number of *Daucus* accessions;

- Characterize the ability of diverse *Daucus* genotypes and derived cell lines to perform SE;

- Develop and establish a collection of *Daucus* cell lines with differential embryogenic efficiencies;

- Collect and characterize basic *AOX* genomic data from the established collection of *Daucus* cell lines;

- Explore *in silico* the capacity of *AOX* polymorphic genes to mark or identify *Daucus* cell lines with differential embryogenic efficiencies;

- Establish new directions for AOX research in the view of marker development concerning *Daucus* SE efficiency.

2

PLANT MATERIAL AND METHODS

2 - Plant material and methods

2.1 - Plant material

Mature *Daucus* mericarps from 28 accessions (Figure 2.1), including 18 *D. carota* cultivars, 5 subspecies and 5 *Daucus* species (see detailed description in the Table 2.1), were used as initial explants to induce *calli* development in an ISE approach. Mericarps were provided by the Julius Kühn-Institut (www.jki.bund.de - Quedlinburg - Germany) (accessions 1 to 27) and by the Institute of Phytopathology and Applied Zoology (www.uni-giessen.de/cms/fbz/fb09/institute/ipaz - Giessen - Germany) (accession 28).

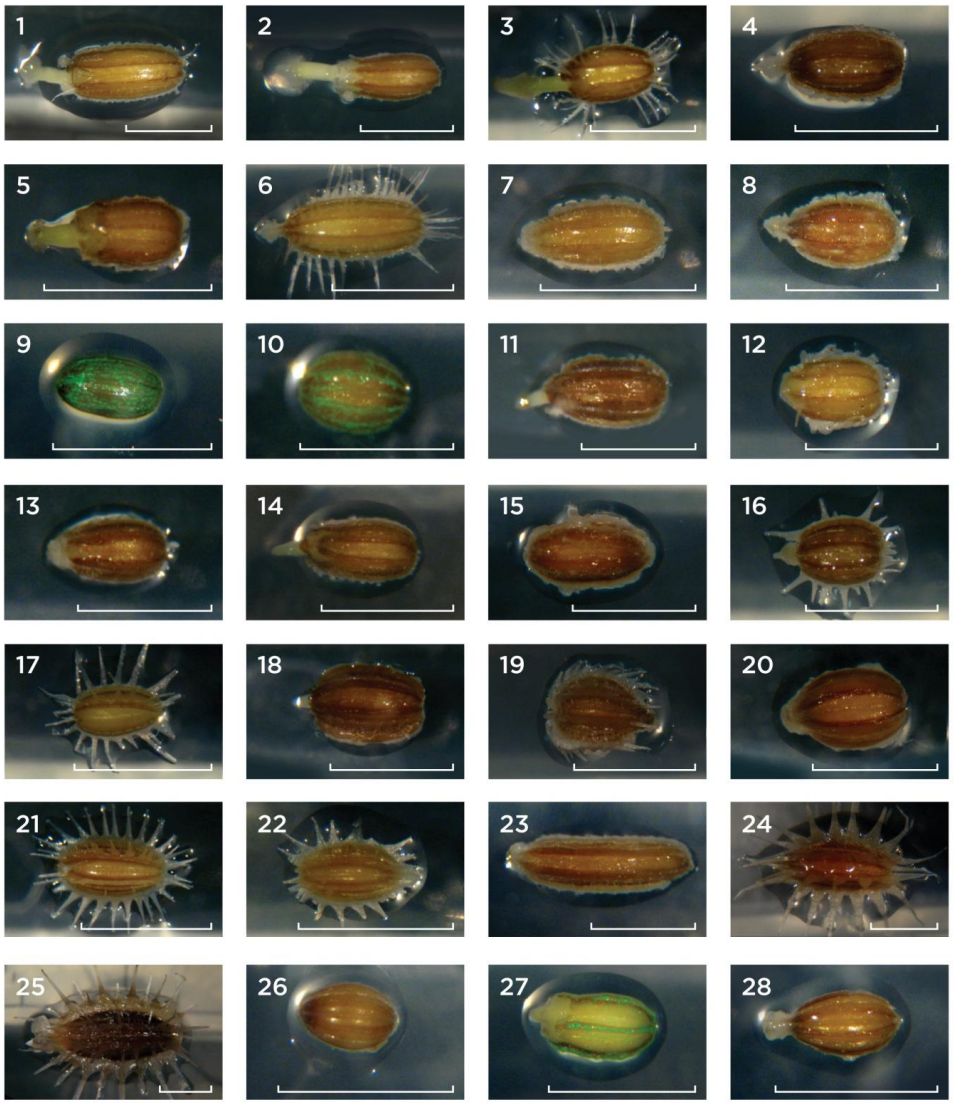


Figure 2.1 - *Daucus* mericarps used as initial explants to induce *calli* development during the ISE approach (for detailed description and accession numbers see Table 2.1). **Bar** - 4 mm.

Table 2.1 - Description and characterization of the 28 *Daucus* accessions used to induce *calli* development during the ISE approach. ***D. c.*** - *Daucus carota*; **Nr.** - Accession number.

		Pedigree (JKI)		Characterization					Origin		
		Code	Generation[1]	Name	2n	Status	Root shape	Phloem	Genebank[2]	Code	Country[3]
Accession (Nr.)	1	LT3-6/01-30	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli*	18	Landrace	Chantenay	Orange	JKI	52136	TUR
	2	LT3-8/01-28	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Senta'	18	Cultivar	Nantes	Orange	IPK	437	DEU
	3	LT3-44/01-21	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes fancy'	18	Cultivar	Nantes	Orange	NGB	1835	DNK
	4	LT1-47/01-10	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes normu'	18	Cultivar	Nantes	Orange	NGB	1856	DNK
	5	LT1-52/01-10	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Amsterdammer master'	18	Cultivar	Amsterdamer	Orange	NGB	1863	DNK
	6	LT3-70/01-30	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nagykallo'	18	Cultivar	Chantenay/Nantes	Orange	HRIGRU	5779	HUN
	7	LT3-100/01-26	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Himuro fuyugosi gosun 2'	18	Cultivar	Flakkeer	Orange	HRIGRU	11718	JPN
	8	LT1-24/01-1	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Lange rote stumpfe 1'	18	Cultivar	Nantes	Orange	IPK	341	DEU
	9	-	OP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Mignon'	18	Cultivar	Nantes	Orange	BEJO Zaden B.V.	-	NLD
	10	-	OP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Vita longa'	18	Cultivar	Flakkeer	Orange	BEJO Zaden B.V.	-	NLD
	11	LT4-59/01-33	SPP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Persia 242'	18	Landrace	Chantenay	Yellow	HRIGRU	3931	IRN
	12	69/01	POP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Western red'	18	Cultivar	Flakkeer	Orange	HRIGRU	5650	AUS
	13	54/01	POP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Norfolk giant'	18	Cultivar	Flakkeer	Orange	HRIGRU	3842	GBR
	14	99/01	POP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Yamanouchi ishyaku senko'	18	Cultivar	Imperator	Orange	HRIGRU	11715	JPN
	15	46/01	POP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes 4 duke'	18	Cultivar	Nantes	Orange	NGB	1855	DNK
	16	DAL 397/00	POP	*D. c.* L. ssp. *commutatus* (Paol.) Thell.	18	Wild	-	White	BGUS	Dau12F/83	ITA
	17	DAL 352/00	POP	*D. c.* L. ssp. *maritimus* (Lam.) Batt.	18	Wild	-	White	JBUL	136/88	PRT
	18	DAL 20/96	POP	*D. c.* L. ssp. *halophilus* (Brot.) A. Pujadas	18	Wild	-	White	JBUL	135/88	PRT
	19	DAL 13/96	POP	*D. c.* L. ssp. *gummifer* (Syme) Hook. f.	18	Wild	-	White	BGUL	19/81	FRA
	20	DAL 17/96	POP	*D. c.* L. ssp. *gadecaei* (Rouy & Camus) Heywood	18	Wild	-	White	IPK	496	FRA
	21	DAL 342/00	POP	*Daucus montevidensis* Link ex Spreng.	22	Wild	-	White	HRIGRU	10459	URY
	22	DAL 340/00	POP	*Daucus pusillus* Michx.	22	Wild	-	White	PSLUR	APE 234	URY
	23	W 85/03	POP	*Daucus capillifolius* Gilli	18	Wild	-	White	HRIGRU	7190	MAR
	24	990/92	POP	*Daucus littoralis* Sibth. & Sm.	20	Wild	-	White	HRIGRU	7997	ISR
	25	DAL 350/00	POP	*Daucus muricatus* (L.) L.	20	Wild	-	White	JBUL	137/88	PRT
	26	-	OP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Rotin'	18	Cultivar	Nantes	Orange	Sperling & Co. GmbH	-	DEU
	27	-	F1	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nevis F1'	18	Cultivar	Nantes	Orange	BEJO Zaden B.V.	-	NLD
	28	-	OP	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Rodelika G280A'	18	Cultivar	Nantes	Orange	Bingenheiner Saatgut AG	-	DEU

1 - SPP - Single plant progeny (seeds were propagated from isolated single plants pollinated by flies - self pollination, genetically these are inbreed lines); **F1** - F1 hybrid cultivar; **OP** - Open pollinated cultivar (produced by breeders in open fields, pollination performed by bees and natural insects); **POP** - Population seeds propagated under isolated conditions (gaze cabins using flies for pollination - genetically these are limited populations because only 5-10 single plants were used as population mother plants); **2 - BEJO Zaden BV** - P.O. Box 50 1749 ZH Warmenhuizen The Netherlands; **BGUL** - Liege University Botanical Garden, Rue Fusch 3, 4000 Liege, France; **BGUS** - Siena University Botanical Garden, Rettorato, Via Banchi di Sotto 55, 53100 Siena, Italy; **Bingenheimer Saatgut AG** - Ökologische Saaten, Kronstrasse 24, D - 61209 Echzell-Bingenheim, Germany; **HRIGRU** - Horticulture Research International - Genetic Resources Unit, Warwick HRI, Wellesbourne, Warwick CV35 9EF, United Kingdom; **IPK** - Leibniz Institute of Plant Genetics and Crop Plant Research, OT Gatersleben, Correns strasse 3, D-06466 Stadt Seeland, Germany; **JBUL** - Lisbon University Botanical Garden, R. da Escola Politécnica 58, Lisboa, Portugal; **JKI** - Julius Kühn - Institut, Federal Research Centre for Cultivated Plants, Erwin-Baur-Str.27, D-06484 Quedlinburg, Germany; **NGB** - Nordic Gene Bank, P.O. Box 41, Alnarp, S - 23053, Sweden; **PSLUR** - Plant Science Laboratories, The University of Reading, Whiteknights, Reading,.Berkshire, RG6 6AS, United Kingdom; **Sperling & Co. GmbH** - Hamburger Straße 35, Lüneburg 21339, Germany; **3 - AUS** - Australia; **DEU** - Deutschland; **DNK** - Denmark; **FRA** - France; **GBR -** Great Britain; **HUN** - Hungary; **ISR -** Israel; **ITA** - Italy; **JPN** - Japan; **MAR** - Morocco; **NLD** - Netherlands; **PRT** - Portugal; **TUR** - Turkey; **URY** - Uruguay.

2.2 - Methods

2.2.1 - Induction of *calli* development and establishment of cell lines

In order to induce *calli* development, mericarps were inoculated, after surface disinfection, in 9 cm diameter (Ø) Petri dishes containing 20 ml of modified Gamborg basal 5 (B_5) (Gamborg *et al*, 1968; Grieb *et al.*, 1997) induction medium, which has been coded as B_5^+. Explants disinfection was performed for 5 minutes with ethanol at 75 % (v/v), followed by an immersion in a solution at 20 % (v/v) commercial bleach with 20 µl of Tween-20 for 20 minutes and subsequently washed twice with sterilized bidistilled water. The medium B_5^+ (1l) was prepared by mixing the following stock solutions volumes (for stock solutions preparation see Table 2.2): 100 ml of macronutrients (10X), 10 ml each of micronutrients (10X), chelated iron (10X) and myo-inositol (10X), 1 ml of vitamins (1000X), 7 ml of magnesium sulfate heptahydrate ($MgSO_4.7H_2O$) (146,06 mM) and 0,5 ml of 2,4-D (4,524 mM). Sucrose grade II (20 g/l) was added directly.

Table 2.2 - List of compounds and their concentrations required to prepare modified B_5 stock solutions. Molarity refers to one liter of stock solution.

Compound name	Formula	Molarity (mM)
Macronutrients stock solution [10X]		
Sodium dihydrogen phosphate	$NaH_2PO_4.2H_2O$	9,62
Potassium nitrate	KNO_3	192,298
Ammonium sulfate	$(NH_4)_2SO_4$	10,141
Magnesium sulfate heptahydrate	$MgSO_4.7H_2O$	20,286
Calcium chloride dihydrate	$CaCl_2.2H_2O$	10,203
Micronutrients stock solution [100X]		
Manganese (II) sulfate monohydrate	$MnSO_4.H_2O$	5,917
Boric acid	H_3BO_3	4,852
Zinc sulfate heptahydrate	$ZnSO_4.7H_2O$	0,174
Sodium molybdate dihydrate	$Na_2MoO_4.2H_2O$	0.058
Copper (II) sulfate pentahydrate	$CuSO_4.5H_2O$	0,010
Potassium iodide	KI	0.452
Cobalt (II) chloride hexahydrate	$CoCl_2.6H_2O$	0,011
Chelated iron stock solution [100X]		
Ethylenediaminetetraacetic acid ferric sodium salt	$C_{10}H_{12}N_2NaFeO_8$	12,614
Vitamins [1000X]		
Nicotinic acid	$C_6H_5NO_2$	4,061
Thiamine hydrochloride	$C_{12}H_{17}ClN_4OS.HCl$	0.296
Pyridoxine hydrochloride	$C_8H_{11}NO_3.HCl$	0,486
Myo-inositol stock solution [100X]		
Myo-inositol	$C_6H_{12}O_6$	14,65

The final volume was adjusted to 1 l after pH correction to 5,72. As gelling agent, agar (6 g/l) was used and added prior to sterilization. Medium was sterilized by autoclaving at 121 °C and 1,05 kg/cm^2 (15 - 20 psi). Sterilization time was dependent upon the volume of medium, varying from 20 (1 l) to 40 minutes (2 l). Sterilized medium was dispensed to sterile Petri dishes on a horizontal laminar flow cabinet. To avoid contaminations gelled medium was coded, sealed and stored in the dark for one week prior use. All chemicals were purchased from Sigma-Aldrich and all were plant tissue culture tested.

Cultures were inspected for contamination on a daily basis. When detecting contamination, the remaining uncontaminated explants were transferred to fresh medium immediately, to avoid dish lost. Cultures were maintained during 6 months, with subcultures of mericarps and seedlings being performed every 14 days to fresh medium. Friable with nodular clumps and pale yellow to white *calli* were collected and separately subcultured into fresh medium. The B_5^+ medium was used throughout all induction subculture steps. The cultures were kept at 25 ºC and under a 16 h light : 8 h dark photoperiod (approximately 35 - 45 µmol m^{-2} s^{-1} of light intensity provided by fluorescent lamps; Philips, Amsterdam, The Netherlands). From each mericarp initially inoculated, a maximum of two *calli* were collected: one developing on the mericarp micropyle region and other developing at the corresponding germinated seedling (seedling: including root, apical meristem and cotyledons). After *calli* isolation, the source tissue was discarded. After 6 months in culture, explants without *calli* development were also discarded.

True-to-type *calli*, i.e. with identical development and appearance as the initially isolated, were maintained and established as cell lines through multiplication by periodic subculture (every 14 days) to fresh B_5^+ medium. *Calli* presenting slow growth, necrosis or changes on appearance were discarded, being considered as not true-to-type.

Established cell lines were identified individually, using a two number coding system. The first number of the code identifies the original material accession number, whereas the second is a sequential number within each accession (see cell line codes in Appendix 1). This coding system is used throughout the text, whenever cell

lines are mentioned. Cell line 5, obtained from a *D. c.* 'Rotin' mericarp (26_L5.S.R), had been the exception to this coding system and was used as reference. Detailed information was gathered for each cell line *calli* concerning original accession source tissue and *calli* appearance in order to check true-to-type cell line development during culture establishment and maintenance (Appendix 1).

2.2.2 - Phenotyping embryogenic efficiency of cell lines

To phenotype each established cell line in terms of SE efficiency, a *calli* portion (not quantified) was transferred individually onto Petri dishes (Ø - 4 cm) containing fresh expression gelled medium, coded as B_5^- and prepared as B_5^+, but lacking 2,4-D (see 2.2.1). SE efficiency was defined as the time (days) required to observe the first somatic embryonic structures (globular, heart or torpedo shaped embryos). For observation and identification of embryonic structures a stereo zoom microscope (Olympus, SZ60 1X-6,3X with a 10X ocular, Japan) was used. Photographic documentation was performed using a digital photographic camera (Canon, PowerShot A630 8.0 mega pixels, Japan) coupled to the microscope through an adaptor (Soligor, adapter tube >52 mm to the Canon PowerShot A630, Israel). *Calli* induced to undergo SE were observed every 15 days, resulting in four observation and documentation points (T15, T30, T45 and T60) during the expression period (60 days) (Figure 2.2). Four plates (considered as replicates) were inoculated per cell line. Two phenotyping sets of four replicates each were performed. Replicates, as well as both phenotyping sets, were performed asynchronously in time.

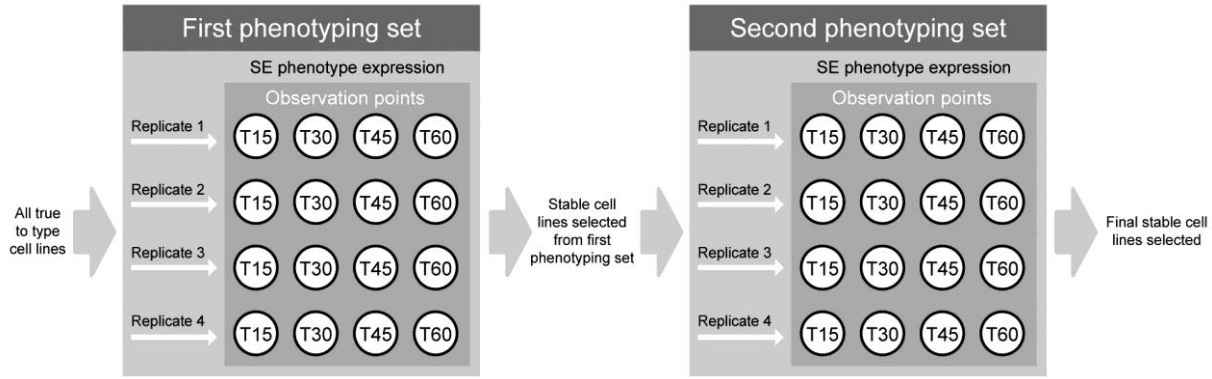


Figure 2.2 - Schematic representation of the experimental design for both phenotyping sets performed to select *Daucus* cell lines with stable SE efficiency phenotype. **Replicates** - 1 to 4 were performed asynchronously in time to improve SE phenotype stability selection over time; **T15, T30, T45 and T60** - Observation and documentation time points, corresponding to 15th, 30th, 45th and 60th day after SE expression start.

For the first phenotyping set all true-to-type cell lines were phenotyped, using *calli* from 9th to 14th day after subculture (composed of cells out of the exponential growth phase). For the second phenotyping set, only the cell lines better scored were selected for re-phenotyping, using *calli* from 6th to 8th day after subculture (during the exponential growth phase).

In the first phenotyping set, the SE efficiency was evaluated regarding the four observation points and the four replicates, being the cell lines classified as very efficient (*VE*, embryonic structures observed at the 15th day), efficient (*E*, 30th day), inefficient (*I*, 45th day), very inefficient (*VI*, 60th day) and non-embryogenic (*NE*, no embryonic structures observed at the 60th day. Cell lines displaying a stable SE efficiency phenotype, i.e. with an identical SE efficiency on three or four replicates of the first set, were selected for genetic diversity evaluation. From those, a subgroup was selected for a second SE efficiency phenotyping set, based on the SE efficiency, stability and genetic diversity evaluation (see section 2.2.3 - Cell lines genetic diversity). The second set was performed similarly to the first, providing a final group of selected cell lines with extreme SE efficiency phenotypes. As extreme SE efficiencies were considered the *VE*, *VE*/*E* and *NE* phenotypes. These final cell lines were maintained as stocks in the undifferentiated state with periodic subculture to fresh B_5^+ every 14 days and used whenever required for further research.

To refine the classification of SE efficiency phenotypes, an additional qualitative scale was created (Appendix 2). This qualitative scale was required because embryonic structures were detected early (at T15 or T30) during replicates, but their number and development varied amongst cell lines, which could lead to erroneous classification and selection. This scale was based on the type of embryonic structures (globular, heart, torpedo or cotyledonary shaped embryos) as proposed by Yeung (1995), but also on the number and the quality of the somatic embryos at the end of the phenotyping period. An effective presence of a higher amount number and better developed embryonic structures on cell lines classified as *VE* in relation to the ones classified as *VI* is clearly documented on the qualitative classification scale to avoid misinterpretation (Appendix 2).

In Appendix 1 all the data are provided concerning SE efficiency phenotyping of the cell lines. Relative frequency of 0,25 was used to mark the occurrence of each SE efficiency phenotype detected per replicate. Cell lines scored with a relative frequency of 1 (0,25 x 4 replicates) were considered as stable, by presenting identical SE efficiency on the four replicates. Cell lines with slight differences in efficiency or with special interest (based on genetic diversity studies) scored with a relative frequency of 0,75 and 0,50 were also considered as minimally stable on the first set. Cell lines scored with a relative frequency of 1 on both sets, were classified as highly stable (eight replicates with identical SE efficiency phenotype).

2.2.3 - Cell lines genetic diversity

Cell lines genetic diversity was evaluated by the cTBP method (combinational tubulin-based polymorphism) described by Breviario *et al*. (2007) and Galasso *et al.* (2011) based on Bardini *et al.* (2004). Genomic DNA (gDNA) was extracted from undifferentiated cell lines *calli* (selected from the first phenotyping set and maintained as stock material) using a DNEasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA integrity was analyzed by electrophoresis on 0,8 % agarose gels (Invitrogen, Thermo Fisher Scientific, MA, USA) after staining in an ethidium bromide solution (0,2 ng/ml) (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). The Gene Ruler™ DNA Ladder Mix (Thermo Fisher Scientific, MA, USA), was used as molecular ruler for band size identification. Gel documentation was performed with the Gene Flash Bio Imaging System (Syngene, Cambridge, UK). gDNA concentration was measured using a NanoDrop-2000C spectrophotometer (Thermo Fisher Scientific, MA, USA). Working solutions (10 ng/µl) were prepared as a template for PCRs.

Exon-primed intron-crossing (EPIC) PCR reaction's were performed using degenerated primers for β-tubulin's introns 1 and 2 ILP fragments amplification (Breviario *et al.*, 2007). PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare, IL, USA) were used following the manufacturer's instructions, adding 10 ng of gDNA and 0.2 µM of each primer. PCR reactions (20 µl) were carried out in a 2720 Thermocycler (Applied Biosystems, Thermo Fisher Scientific, MA, USA) following described parameters (Breviario *et al*., 2007). PCR control reactions with no primers

or single primers were performed for each cell line. Three PCR reactions were performed per cell line. PCR products (8 µl) were separated by vertical electrophoresis in polyacrylamide gels (6 % w/v) for 4 h at 100 V. Two electrophoresis runs were performed per PCR, totalling six per cell line. Gel staining and visualization was performed as described above. All six gels performed for each cell line revealed identical pattern of amplified β-tubulin's ILP's. The pattern of amplified fragments (markers) between cell lines was used to construct a matrix of presence (1) / absence (0). ILP markers were individually identified using arabic numbers sequentially. Numbering started from intron 1 to intron 2 and from the lowest molecular size marker to the highest. The FreeTree software (Pavlicek *et al.*, 1999; Hampl *et al*., 2001) was used to compute the distance/similarity matrix according to Nei and Li (1979), as well to construct the Unweighted Pair Group Method with Arithmetic Average (UPGMA) (Sokal and Michener, 1958; Murtagh, 1984) dendrogram and to perform the bootstrapping (Efron, 1979) with 1000 replicates.

2.2.4 - Cell lines ploidy

For ploidy determination, flow cytometry was performed as an external service provided by the Plant Cytometry Services Company (www.plantcytometry.nl, Netherlands). Plants from each of the accessions selected at the second phenotyping set end, were germinated from mericarps under controlled conditions (see section 2.2.1). They were the source of leaves, which had been used as 2n=2x=18 control for flow cytometry measurements. Cell lines *calli* and leaf controls were analysed with an internal standard (*Buxus sempervirens* L.) using 4',6-diamidino-2-phenylindole (DAPI) as fluorescent dye. Ploidy levels were expressed as the DNA ratio measured by the internal standard among the leaves used as control and the corresponding cell line *calli*. The n value was determined by the relation between the cell line and its control leaf DNA content.

2.2.5 - Amplification of *AOX1, 2a and 2b* in cell lines

Cell lines *calli* were used for DNA extraction which was performed using the DNeasy Plant mini kit (Qiagen, Hilden, Germany) according to the manufacture's instructions

and quantified using the Nanodrop 2000C (Thermo Fisher Scientific, MA, USA). Working solutions (10 ng/µl) were used as template. PCR reactions (50 µl) were carried out in a 2720 Thermocycler (Applied Biosystems, Thermo Fisher Scientific, MA, USA) using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, MA, USA), and specific primers to each of the amplified *AOX* (Table 2.3). The PCR program consisted of an initial step at 98 ºC during 30 seconds for initial template denaturation, followed by 35 cycles each consisting of: 98 ºC during 20 seconds for denaturation; 64, 52 or 55 ºC (*AOX1*, *AOX2a* and *2b*, respectively) during 30 seconds for primers annealing; and 72 ºC during 2 minutes for chain extension. After the last cycle, a final extension step was performed during 10 minutes at 72 ºC.

PCR products were separated by electrophoresis in a 1,4 % agarose gel (Invitrogen, Thermo Fisher Scientific, MA, USA) during 60 minutes at 100 V, and stained during 35 minutes in an ethidium bromide solution (0,2 ng/ml) (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). The bands of interest were excised from the agarose gel using a sterile scalpel blade under long wave ultraviolet (UV) light (365 nm) to avoid DNA damage, provided by a dual UV transilluminator (VWR, PA, USA). DNA was recovered from agarose using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, IL, USA) and quantified using known standard λDNA samples (Fermentas, Thermo Fisher Scientific, MA, USA). After quantification, recovered DNA was inserted in the pGem®-T Easy Vector (Promega, WI, USA) and incubated overnight at 4 ºC to maximize retrieval of recombinants. The vector with the insert was used to transform *Escherichia coli* competent cells JM109 (Promega, WI, USA).

Table 2.3 - Characterization of primers used for *Daucus AOX* genes amplification. **Dc** - *Daucus carota*; **Fw, FW** - Forward; **Number** - Primer position in the sequence; **Rv**, **Rev** - Reverse; **UTR** - Untranslated region.

Gene	Name	5' - 3'
AOX1	DcAOX1a24Fw	TGA AAA TAA CAA TGA TGA TGA C
	DcAOX1a1032Rv	AAC CAG AGA TTC CTC CAC TTC A
AOX2a	DcAOX2a30Fw	ATG AAT CAT CTG TTA GCC AAG TCT G
	DcAOX2a_3'UTR	TTC AGA GAT ATA TAG CTA TGT GG
AOX2b	DcAOX2b_40FW	TGC ATG CGT CCT TCC TTA TTT TTC
	DAOX2b_1188Rev	GCT CTG CTG TGA TTT TCT GGA C

The transformed competent cells were grown overnight using selective Luria Bertani (LB) Agar High Salt (Duchefa Biochemie, Haarlem, The Netherlands), supplemented

with 500 µl of Carbenicillin disodium (100 mg/ml) (Duchefa Biochemie, Haarlem, The Netherlands), 500 µl of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0,5 M) (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany) and 800 µl of 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) (50 mg/ml) (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). Recombinant clones were picked from plates and re-grown overnight in LB Broth High Salt medium (Duchefa Biochemie, Haarlem, The Netherlands). For *AOX2a*, all above steps were performed using the TOPO XL PCR Cloning Kit (Invitrogen, Thermo Fisher Scientific, MA, USA) especially designed for cloning of long PCR products (3 - 10 kb), according to the manufacturer's instructions. Plasmid DNA was recovered by alkaline lysis as proposed by Bimboim and Doly (1979). After recovery, plasmid DNA was digested with *EcoR*I restriction enzyme (Thermo Fisher Scientific, MA, USA). The digestion products were analysed and bacterial clones presenting the fragments of interest selected for sequencing.

2.2.6 - Sequencing and bioinformatic analyses

Plasmid DNA was quantified similarly to gDNA and sequenced using the universal primers T7, SP6 and M13R-pUC. Sequencing was performed as an external service at the Macrogem Company (www.macrogen.com). Additional internal primers were used whenever required (Table 2.4).

Table 2.4 - Characterization of primers used for *AOX* genes sequencing. **Dc** - *Daucus carota*; **Fw** - Forward; **G** - Genomic; **int** - internal; **Number** - Primer position in the sequence; **R**, **Rev** - Reverse.

Gene	Primer name	5' - 3'
AOX1	DcAOX1aR	ATC TCG CAA TGT AGA GTC AGC C
*AOX2*a	DcAOX2a_4676GFw	CTC TGT TTC ATA TTA CAT GTC C
	DcAOX2a_929GRev	GCA GAG TCA GAT CCA ATT TAT G
	DcAOX2a_900GRev	AGT CTG ATA CCA TAT TAT AGG
	DcAOXFw2	TGA GGT GTG TAT ATT TTT TGC
	DcAOX2a_Rev2	GCT CAT CCA CGC GCA CTC T
	DcAOX2a_Rev3	GGA GTT TTT GAA TGC TGA TA
	DcAOXFw3	AGA GTA GCT AAT TAG TGT GG
	DcAOX2a_Rev1	GGA GTT GGT TAT ATCGT
	DcAOX2a_852Fw	CAA TTG AAA ATG TTC CTG CTC C
	DcAOX2a_int2R	TAA GCA CCA TGT ACC AAA GAC
AOX2b	DcAOX2bR	CGT ATA ACT AGT ATA ACA TCT CTC
	DcAOX2b_1255Rev	TAT TCA GAT CAA TGG ACA CG

Obtained sequences were trimmed using the EditSeq application, of the software Lasergene suite V7.1.0 (DNASTAR, WI, USA). Whenever required, re-sequencing in the forward and reverse strand was performed to confirm polymorphic positions or unique sequences. Trimmed sequences were merged on the application SeqMan, also from the Lasergene suite V7.1.0 software (DNASTAR, WI, USA), in order to obtain full length *AOX* sequences.

Merged *AOX* sequences were blasted at the National Center for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov/blast) to confirm *AOX* identity, applying the algorithm Basic Local Alignment Search Tool (BLAST - BLASTx and BLASTn) (Karlin and Altschul, 1993). Open reading frames (OFRs) were deduced and confirmed using the GeneMark webserver (Borodovsky and McIninch, 1993) and by BLAST at NCBI database. Amino acid sequences were obtained using EditSeq, also from the Lasergene suite V7.1.0 software (DNASTAR, WI, USA). To obtain an evolutionary history and identify possible polymorphic sites of interest, precise alignments were made for each set of *AOX* sequences (nucleotide and amino acid). To perform global and coding region alignments, FASTA format files were created and used in the software Alignment Sequence Editor BioEdit V7.0.9.0 (Hall, 1999) to run the Clustal W alignment method (Thompson *et al.*, 1994; Larkin *et al.,* 2007). For regions with a high degree of variation (introns) the multiple alignment using fast Fourier transform (MAFFT) method was used instead (mafft.cbrc.jp/alignment) (Katoh *et al.*, 2002; Katoh *et al.*, 2005). The iterative strategy E-INS-I (very slow; recommended for < 200 sequences with multiple conserved domains and long gaps) was applied using the standard parameters. Manual editions were performed whenever required, to correct specific alignment positions. CLC Workbench 6.9.2 software was used to perform all annotations and capture alignments images.

Phylogenetic and molecular evolutionary analyses were conducted with MEGA version 5 (Tamura *et al.*, 2011) using the Neighbor-Joining method (Saitou and Nei, 1987) and the phylogeny test Bootstrap (1000 replicates) (Felsenstein, 1985).

2.2.7 - *In silico* analysis of *AOX*: Functional sites, subcellular localization and regulatory elements

The sequence harmony (SH) method, as previously described by Pirovano *et al*. (2006), and made freely available on the web (http://www.ibi.vu.nl/programs/) (Feenstra *et al*., 2007), was used to perform AOX amino acid sequence comparison, in order to identify specific putative functional sites using a cutoff of 0,1 as comparison parameter. Sequences were read from the alignment and after been separated into two user-specified groups (AOX1 and AOX2). For each group individually and combined entropies were calculated. SH score values can range from zero for completely non-overlapping residue compositions, to one for identical compositions. Residue positions were selected based on the SH scores below the cutoff. Stretches of neighboring selected positions were identified and the size of each of these stretches was assigned to the positions as the rank. As a result selected residue positions were sorted as: increasing SH, decreasing rank and increasing entropy. This sorted list of selected residue positions was the primary result of the SH algorithm, which had been transformed and sorted by alignment residue position for better visualization and interpretation.

TargetP 1.1 web server (freely available at www.cbs.dtu.dk/services/TargetP/) was used to determine targeting peptide (TP) scores and potential cleavage sites using default parameters for plants (Nielsen *et al.*, 1997; Emanuelsson *et al.*, 2000) in order to predict the subcellular localization of identified AOX protein variants.

IMEter v2.1 (http://korflab.ucdavis.edu/cgi-bin/IMEter_2014/web-imeter2.1.pl) (Rose *et al.*, 2008; Parra *et al.*, 2011) was used to infer the intron-mediated enhancement (IME) scores in the forward and reverse strand according to default parameters. *A. thaliana* was the species selected as reference in the software to perform the comparison with the amplified *Daucus* introns under evaluation.

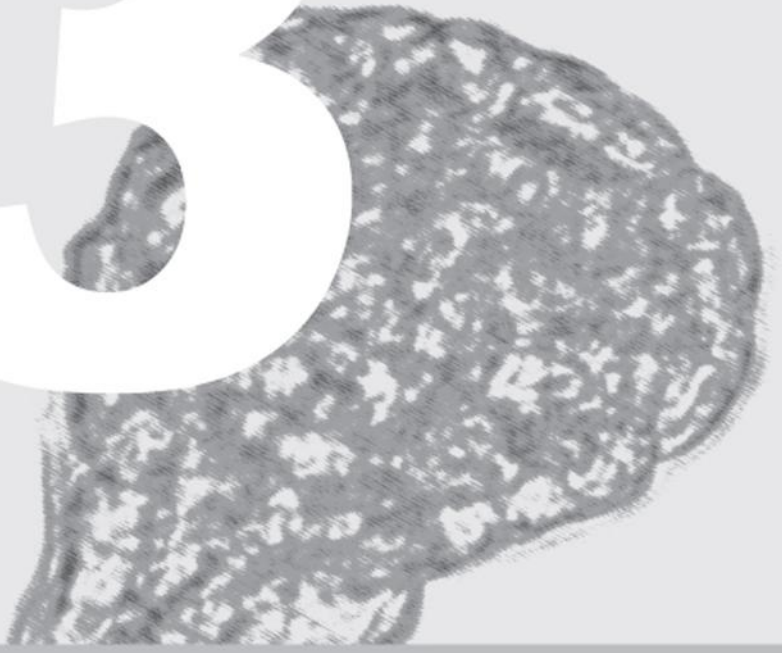
The software Nsite-PL Version 5.2013 (Shahmuradov *et al.*, 1986; Solovyev and Kolchanov, 1994; Solovyev, 2002; Shahmuradov and Solovyev, 2015) and ScanWM-P (both freely available at the Softberry Inc. web portal - www.softberry.com), were used for search of *cis*-acting regulatory elements in the unaligned *Daucus AOX*

highly polymorphic regions (ORFs and intron 1). The RegSite database of regulatory elements containing 2779 plant transcription regulatory sites (updated in May, 2014 with elements retrieved from published data on transcription regulation of plant genes) was employed as the elements reference database for both software during the search.

Nsite-PL performed the search based on statistically significant functional motifs of plant regulatory sequences, using the statistical estimation of the expected number of a nucleotide consensus pattern in a given sequence. ScanWM-PL performed the search for functional motifs described by weight matrixes of plant regulatory sequences, which was built from a subset of plant regulatory sequences from RegSite Database. Both software performed the search along the forward and reverse strand. For the software Nsite-PL, as search parameters were used, 0,1 as expected mean number, 99 % as statistical significance level, 100 % as level of homology between known regulatory element and motif, and a variation of distance between regulatory elements blocks of 20 %. In the ScanWM-P the search parameters used were, threshold type of 2 and threshold value of 99 %.

Regulatory elements found at polymorphic positions/regions (SNPs or InDels) were seen as having a putative differential functional importance, and used for comparison amongst identified sequences with identical SE phenotype, throughout the construction of tables displaying the elements found, but in the aligned position. Appendix 9 displays the constructed tables, showing discovered elements at polymorphic positions, and its placement in the aligned sequences. Elements found to be present in all the sequences were not considered relevant for the analyses and were not represented in the tables. Elements discovered in a subset of sequences with equal phenotype were considered as possibly functionally important. Whenever relevant, elements were reviewed in literature for further information concerning SE.

3



RESULTS

3 - Results

3.1 - Mericarp germination, *calli* development and cell line establishment

Daucus mature mericarps were inoculated, after disinfection, on B_5^+ medium supplemented with 2,4-D (see section 2.2.1 and Frederico *et al*., 2009a), to initiate germination under identical conditions.

Fourteen days after *in vitro* inoculation, when the first subculture to fresh culture medium was performed, most of the mericarps presented already some visible morphological changes (enlargement), attesting their capacity to respond under the tested conditions. Germination was highly dependent on the accession and ranged from ten days in most *D. carota* cultivars, up to four months for *D. montevidensis* (accession 21). In two accessions, 24 (*D. litorallis*) and 25 (*D. muricatus*), both not *D. carota*, no germination was observed by the end of the experiment (Table 3.1). Nevertheless, mericarps from these accessions were able to germinate under identical culture conditions, in other experiments, by using a medium devoid of 2,4-D (data not shown). Until 30 days after inoculation, the overall germination rate was 67,81 % (Table 3.1). On average, the highest germination rates were obtained with the *D. carota* cultivars followed by *D. carota* subspecies, 83,50 % and 43,85 %, respectively. Only a few *D. carota* accessions showed delayed germination in some of the mericarps. These were the cases of accessions 7 (*D. carota ssp. sativus* 'Himuro fuyugosi gosun 2'), 11 (*D. carota ssp. sativus* 'Persia 242') and 20 (*D. carota ssp. gadecaei*) (Table 3.1; column: Germinated, with the mark *). For accession 22 (*D. pusillus*), mericarp germination was only observed 30 days after *in vitro* inoculation, being therefore the most recalcitrant to germinate.

After mericarp germination, the tissues in culture, including seedlings and mericarp coat remains, were maintained and subcultured (Figure 3.1- B, E and F). All were considered as a tissue source for the development *of calli*. The development of *calli* was visible in the tissues along the first four months of the induction period, depending on the accession, germination occurrence and tissue source. The capacity to develop *calli* was directly correlated with the mericarp capacity to germinate. The exception was accession 22 (*D. pusillus*), which presented germinated mericarps but did not develop *calli*.

Table 3.1 - Global overview of the experiments performed to induce *calli* development from the 28 *Daucus* accessions evaluated (see Table 2.1 for accession characterization). Data represents the number (Nr.) of mericarps inoculated per accession, the number of germinated mericarps and the germination percentage (%). Also shown are the number of true-to-type cell lines isolated by genotype and the number of cell lines used per SE phenotyping set (1st and 2nd). Final number of cell lines selected is also indicated (see section 3.2). ***D. c.*** - *Daucus carota*; ***** - Mericarps germinated after 30 days.

			Mericarps (Nr.)			Cell lines (Nr.)			
		Name	Inoculated	Germinated	%	Isolated	1st set	2nd set	Selected
Accession (Nr.)	1	*D. c.* ssp. *sativus (Hoffm.) Arcangeli*	7	6	86	9	7	1	0
	2	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Senta'	8	6	75	8	8	1	1
	3	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes fancy'	9	9	100	8	8	3	0
	4	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes normu'	11	6	55	11	8	2	0
	5	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Amsterdammer master'	10	4	40	9	8	1	1
	6	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nagykallo'	10	6	60	6	6	2	0
	7	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Himuro fuyugosi gosun 2'	9	8 (1)*	89	10	8	1	0
	8	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Lange rote stumpfe 1'	8	4	50	6	5	1	1
	9	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Mignon'	10	5	50	7	7	0	0
	10	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Vita longa'	9	9	100	8	7	1	0
	11	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Persia 242'	10	8 (1)*	80	6	4	0	0
	12	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Western red'	10	8	80	11	4	0	0
	13	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Norfolk giant'	12	9	75	9	4	1	0
	14	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Yamanouchi ishyaku senko'	13	6	46	11	4	0	0
	15	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nantes 4 duke'	11	9	82	7	4	0	0
	16	*D. c.* L. ssp. *commutatus* (Paol.) Thell.	12	1 (3)*	8	3	3	0	0
	17	*D. c.* L. ssp. *maritimus* (Lam.) Batt.	12	9	75	9	6	1	0
	18	*D. c.* L. ssp. *halophilus* (Brot.) A. Pujadas	12	7	58	9	6	2	1
	19	*D. c.* L. ssp. *gummifer* (Syme) Hook. f.	11	6 (2) *	55	6	4	1	1
	20	*D. c.* L. ssp. *gadecaei* (Rouy & Camus) Heywood	10	2 (1)*	20	3	3	1	1
	21	*Daucus montevidensis* Link ex Spreng.	11	2 (8)*	18	2	2	0	0
	22	*Daucus pusillus* Michx.	10	0 (5)*	0	0	0	0	0
	23	*Daucus capillifolius* Gilli	10	1 (9) *	10	6	6	0	0
	24	*Daucus littoralis* Sibth. & Sm.	10	0	0	0	0	0	0
	25	*Daucus muricatus* (L.) L.	12	0	0	0	0	0	0
	26	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Rotin'	50	50	100	18	6	1	1
	27	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Nevis F1'	50	45	90	31	5	1	1
	28	*D. c.* ssp. *sativus (Hoffm.) Arcangeli* 'Rodelika G280A'	50	50	100	24	6	1	0
		Total	**407**	**276 (30)***		**237**	**139**	**22**	**8**

Calli with non-embryogenic and embryogenic characteristics, were identified in the cultures during the induction period. The typical non-embryogenic *calli* (mucilaginous and translucent, strongly green or dark in color) were mainly present in seedling tissues (Figure 3.1 - A, B, C and D), from where later, some typical embryogenic *calli* (white, pale yellow or pale green and friable with nodular clumps) arises, reason why it was impossible to observe from each seedling tissue typical embryogenic *calli* were derived. *Calli* with typical embryogenic characteristics (friable with nodular clumps, with a color range from pale yellow to white and with a fast proliferation rate), occurred also in the mericarp coat remains at the seedling emergence site (Figure

3.1 - E and F), being accession-dependent (accessions 1, 2, 3, 4, 5, 7, 8, 12, 14, 23, 26, and 28) (Appendix 1).

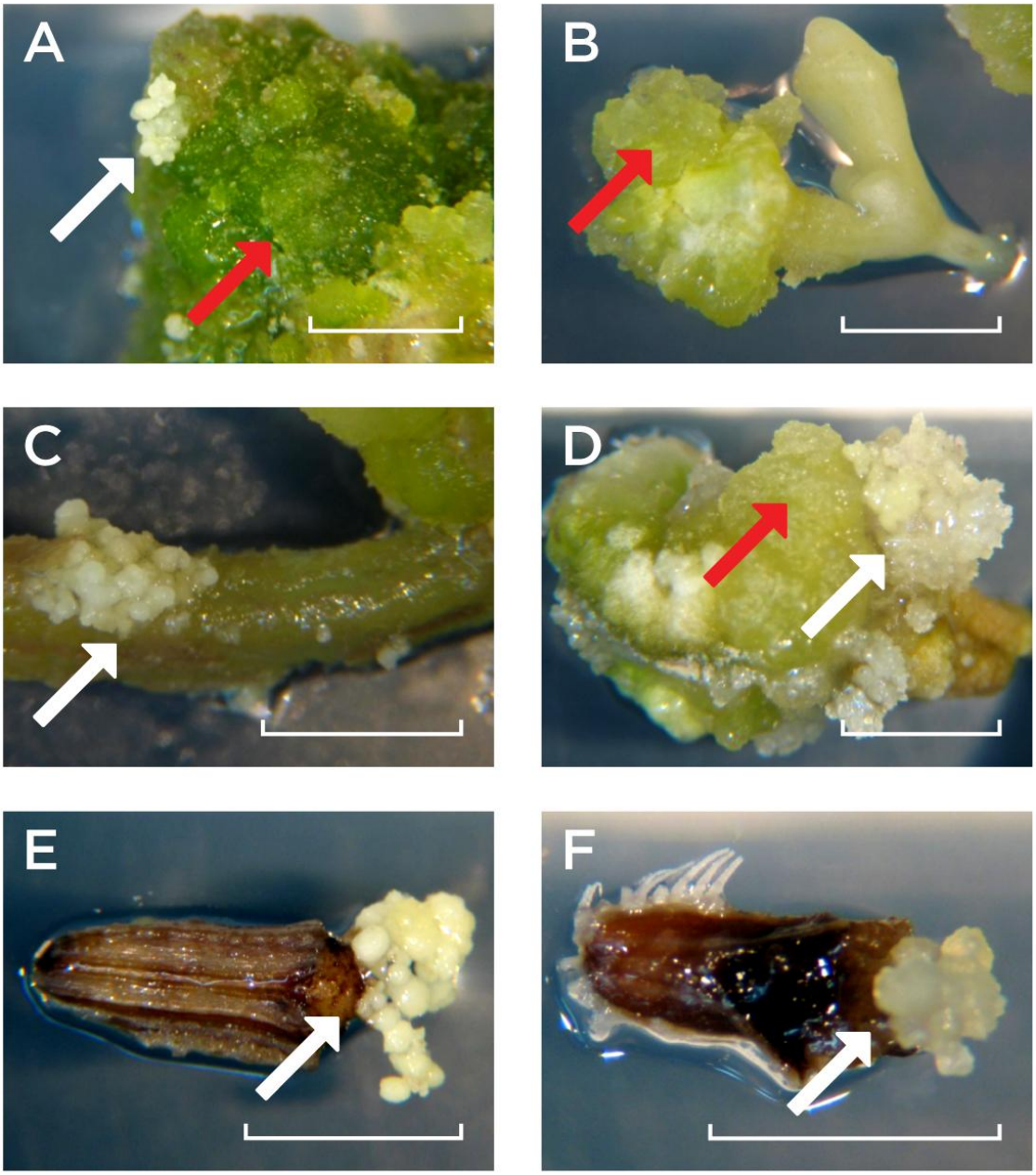


Figure 3.1 - *Calli* development on seedling tissues (A, B, C and D) and mericarp internal tissues (E and F) during the induction phase. **Arrows** - *Calli* with typical embryogenic (white) or non-embryogenic (red) characteristics in *Daucus*; **Bar** - 4 mm

From the 25 accessions which developed *calli*, 237 with a typical embryogenic appearance were isolated from seedling and mericarp tissues (Figure 3.1), being individually re-conducted in culture (Table 3.1). However, variability in terms of *calli* appearance and growth performance after isolation has been observed, namely: degeneration, necrosis and slow growth.

Aiming at maintaining only true-to-type cultures, 98 (41,4 %) *calli* with visible changes in growth or appearance were discarded. From the true-to-type *calli*, 139 (58,6 %), cell lines were successfully established (Table 3.1), without any further change, being 113 (81,3 %) derived from seedling tissues and 26 (18,7 %) from mericarp internal tissues (for details see Appendix 1). In general, true-to-type *calli* were friable with nodular clumps, firm or viscous and with four color variants, yellow, greenish, white or brown/dark. The final number of cell lines established per accession is presented in Table 3.1. The individual characterization for each one of those cell lines is presented in Appendix 1 and reported in detail in the following sections.

3.2 - Phenotyping cell line SE efficiency

In accordance to Yeung (1995), SE efficiency was set as the time (days) required to observe the first embryonic structures. To induce SE expression and evaluate its efficiency, *calli* portions were transferred from induction B_5 medium supplemented with 2,4-D (B_5^+) to fresh expression B_5 medium without 2,4-D (B_5^-). Preliminary evaluations were performed to establishing the optimal time range required to cover all SE efficiency variants among the 139 true-to-type established cell lines. These evaluations showed that no further development of embryonic structures was observed after 60 days following SE induction (data not shown). On these preliminary evaluations, the need to perform regular subcultures during the SE experiments was also assessed. Nevertheless, similar outcomes were obtained when experiments were performed with or without a subculture (data not shown).

Based on this preliminary evaluation data, phenotyping observations and documentation were established to be performed every 15 days during the 60 day's period of the SE expression experiments. As mentioned in section 2.2.2, four observation time points were performed for each phenotyping replicate, at the 15^{th} (T15), 30^{th} (T30), 45^{th} (T45) and 60^{th} (T60) days after cell line SE expression start. The inoculation time point (T0) was also documented, as well as a final overview of cultures. Based on the results achieved from the two phenotyping sets, each comprising four replicates and the four observations per replicate, the cell lines were classified as very efficient (*VE*, embryonic structures observed at T15), efficient (*E*, embryonic structures observed at T30), inefficient (*I*, embryonic structures observed

at T45), very inefficient (*VI*, embryonic structures observed at T60), or non-embryogenic (*NE*, no embryonic structures observed at T60).
To support cell lines selection and improve its classification, an additional qualitative scale was created (Appendix 2), due to the fact that embryonic structures were detected early in the replicates (at T15 or T30), but their stage of development varied amongst cell lines, which could lead to erroneous selection. The scale is separated in seven quality categories of embryogenic cell lines, coded using the signals - and +, and one category of non-embryogenic cell lines, coded using the signal #. This scale was based on the observed number of embryonic structures and their developmental stage (globular, heart, torpedo, and cotyledonary shaped embryos), and was used to rank cell lines at the end of each phenotyping experiment set. Cell lines coded with ---, -- or - have low amount of embryogenic structures and low quality embryos. On the other hand, cell lines coded with +++, ++ or + have a high amount of embryogenic structures and high quality embryos.

Appendix 1 presents data collected at the observation points (T15, T30, T45 and T60) performed during both SE phenotyping sets. Data includes phenotype relative frequency and stability and also the amount and quality of embryonic structures. Jointly, this data was applied to classify cell lines and perform a selection for extreme embryogenic behaviors, which will be reported in detail in the next sections (see section 3.2.1 and 3.2.3).

3.2.1 - First phenotyping set for SE efficiency

Induction of SE is usually performed during *calli* exponential growth, which occurs in *Daucus* cells, between the 6^{th} and 8^{th} days after subculture (see Frederico *et al*., 2009a). The use of *calli* out of the exponential growth phase (from 9 to 14 days after subculture), was expected to be an additional selection factor facilitating the identification of cell lines with higher efficiency stability amongst replicates.
Concerning the first set, 139 cell lines from 25 accessions were phenotyped. In total, 556 phenotyping experiments (139 cell lines x 4 replicates) were conducted with four observation time points (T15, T30, T45 and T60), resulting in 2224 phenotyping observations (556 phenotyping experiments x 4 observation time points) performed and documented.

Organized embryonic structures (globular, heart and torpedo embryos) were detected at the first observation performed 15 days after transfer of *calli* to culture medium devoid of 2,4-D (B_5^-), as well as in all the other three time points (T30, T45 and T60). Under the tested conditions all the 25 accessions were able to differentiate somatic embryos, at least, in one cell line. Six (24 %) were represented exclusively by embryogenic cell lines. The remaining 19 (76 %) were detected, at least in one cell line or in one of four replicates performed, without embryogenic capacity.

Figure 3.2 and Appendix 1 provide the detailed data on the SE efficiency phenotype analyses of individual accession cell lines four replicates in the first phenotyping set. For each detected SE efficiency phenotype (*VE*, *E*, *I*, *VI* and *NE*), the relative frequency of 0,25 was used to identify the phenotype occurrence (see Appendix 1). Concerning the *NE* phenotype 53 cell lines express it in 117 phenotyping experiments. Of those, 11 express this behavior in a stable manner in all the four replicates. Remaining 42 cell lines proved to be unstable, by not expressing the NE phenotype in at least one of the replicates.

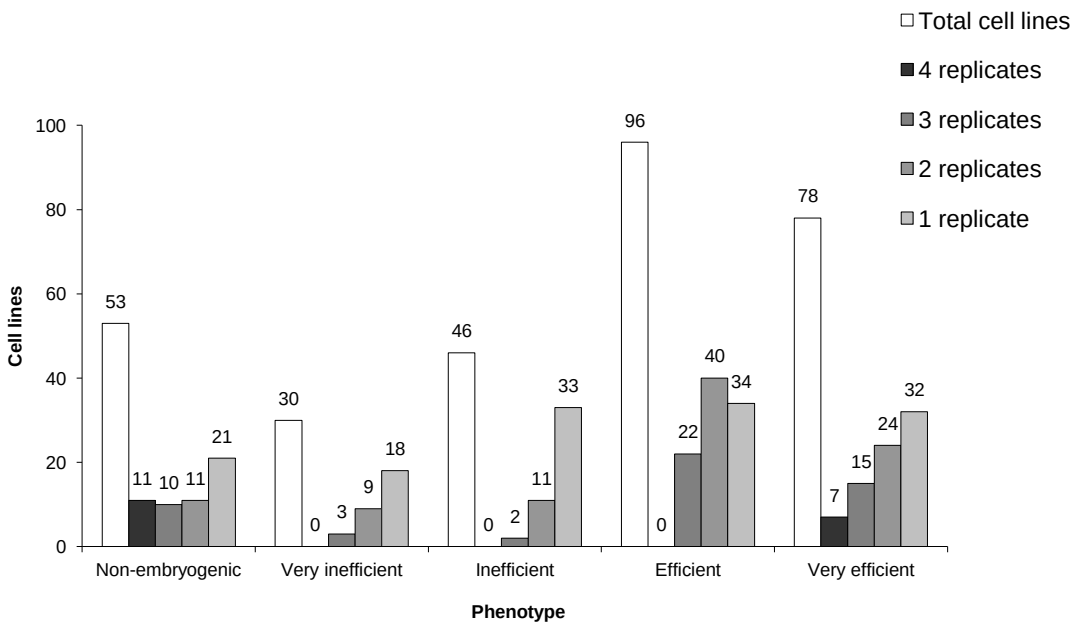


Figure 3.2 - Graphical representation of cell line SE efficiency phenotypes obtained from the first phenotyping experiment set. Phenotyping was carried out on gelled medium B_5^- during a 60 day period and evaluation and documentation were performed at four different time points (T15, T30, T45 and T60) after SE induction. The original data used for generating the graph are provided in Appendix 1. **Grey bars** - Total number of cell lines showing an embryogenic phenotype in four (dark grey), three (mid-dark-grey), two (mid-light-grey) or one (light grey) replicates; **White bars** - Total number of cell lines detected per embryogenic phenotype, independently of the number of replicates.

For the *VI*, *I* and *E* phenotypes, 30, 46 and 96 cell lines express it in 45, 61 and 180 phenotyping experiments respectively, all in an unstable manner. None of these cell lines were capable to express the phenotypes consistently in all the four replicates.
For the *VE* phenotype, 78 cell lines express it in 153 phenotyping experiments. From those, seven expressed the phenotype in a stable manner. The remaining 71 expressed the phenotype in an unstable manner, in at least one of the replicates.
After the analyses of the first phenotyping set data (Figure 3.2), it was possible to classify all defined embryogenic phenotypes at least in one of the replicates. From 139 cell lines, 98 (70,5 %) expressed embryogenic efficiency phenotype in an unstable way, with clear differences among replicates, being thus eliminated. The other 41 (29,5 %) showed an evident efficiency stability. Nonetheless, only 18 (43,9 %) expressed it equally in all four replicates. The remaining 23 (53,1 %) showed a weaker efficiency stability, by showing adjacent efficiencies in one or two out of four replicates, but never being divergent. In this sense and due to the existence of uncertainties about its efficiency stability, these 23 cell lines were considered for further evaluation. All 41 cell lines considered as minimally stable at the end of the first phenotyping experiment set were kept as stock material and their genetic diversity was evaluated (see section 2.2.3) using the cTBP method (see section 3.2.2).

3.2.2 - Cell lines genetic diversity

To assess, evaluate and characterize the genetic diversity background of the 41 cell lines identified as holding a stable efficiency phenotype and also to search for possible correlations at the SE efficiency, source tissue or phylogenetic level, a molecular marker was applied. The selected molecular marker was the one that uses intron 1 and 2 polymorphisms of the β-tubulin gene family, through a combinatorial tubulin-based polymorphism method (cTBP) (Breviario *et al*., 2007). Both, intron 1 and 2 polymorphic ILP fragments were used in combination due to the higher number of markers produced when compared with the single use of intron 1 ILP's (Breviario, 2007). This combined method was more appropriate to characterize individual differences within accessions, as in the case of multiple cell lines obtained from a single accession.

The cTBP gel band patterns obtained for the 41 cell lines evaluated, are presented in Figure 3. It had been used to detect bands presence or absence, in order to assess the individual cell lines cTBP ILP patterns. Appendix 4 contains the ILP data pattern detected from each cell line, represented as the presence (1) or absence (0), as well as the individual band number identification per intron evaluated. Appendix 5 shows the similarity matrix obtained as mentioned in the section 2.2.3, which had been applied to construct the dendrogram (Figure 3.4), showing the genetic distance amongst the cell lines evaluated.

Considering both introns, 75 bands were identified (73 polymorphic and two monomorphic). Intron 1 was represented by 43 polymorphic bands and intron 2 by 32 bands, two of them being monomorphic (band 45 and 55, respectively) (Appendix 4). Also considering both introns, the cell line with the highest number of bands (34 bands) identified was 27_2 (*D. c.* 'Nevis F1'). On the other hand, the cell line 15_4 (*D. c.* 'Nantes duke') with 16 bands identified, was the one with the lowest cTBP polymorphism. The dendrogram representing the genetic diversity obtained from the cTBP marker assessment is shown in the Figure 3.4. The cTBP markers were able to detected differences in all the cell lines evaluated, with the exceptions of those from cultivars 'Nantes normu' (4_5 and 4_1), 'Amsterdammer master' (5_4 and 5_7) and 'Nantes fancy' (3_5 and 3_7). In general, cTBP markers were able to identify and clearly separate the *Daucus* cultivars from the subspecies and species. However, some cell lines derived from *D. c.* cultivars, cluster with cell lines from other cultivars, as in the case of the *D. c.* 'Senta', which has cell lines in several clusters.

The 22 cell lines, identified with arrows in Figure 3.4, were selected for re-evaluation of the SE phenotype in a second phenotyping set (see section 3.2.3). Those were chosen based on the SE phenotype stability, identified during the first phenotyping set and the goal was to maximize the coverage of the genetic diversity under study. The two groups of cell lines with identical cTBP patterns (4_1 / 4_5 and 3_6 / 3_7) had been selected, because they had different SE phenotypes in the first phenotyping set. Identical situation happened when multiple cell lines from a single accession had been selected, as in the case of the accessions 6 (*D. c.* ‘Nagykallo’) and 18 (*D. c*. *halophilus*).

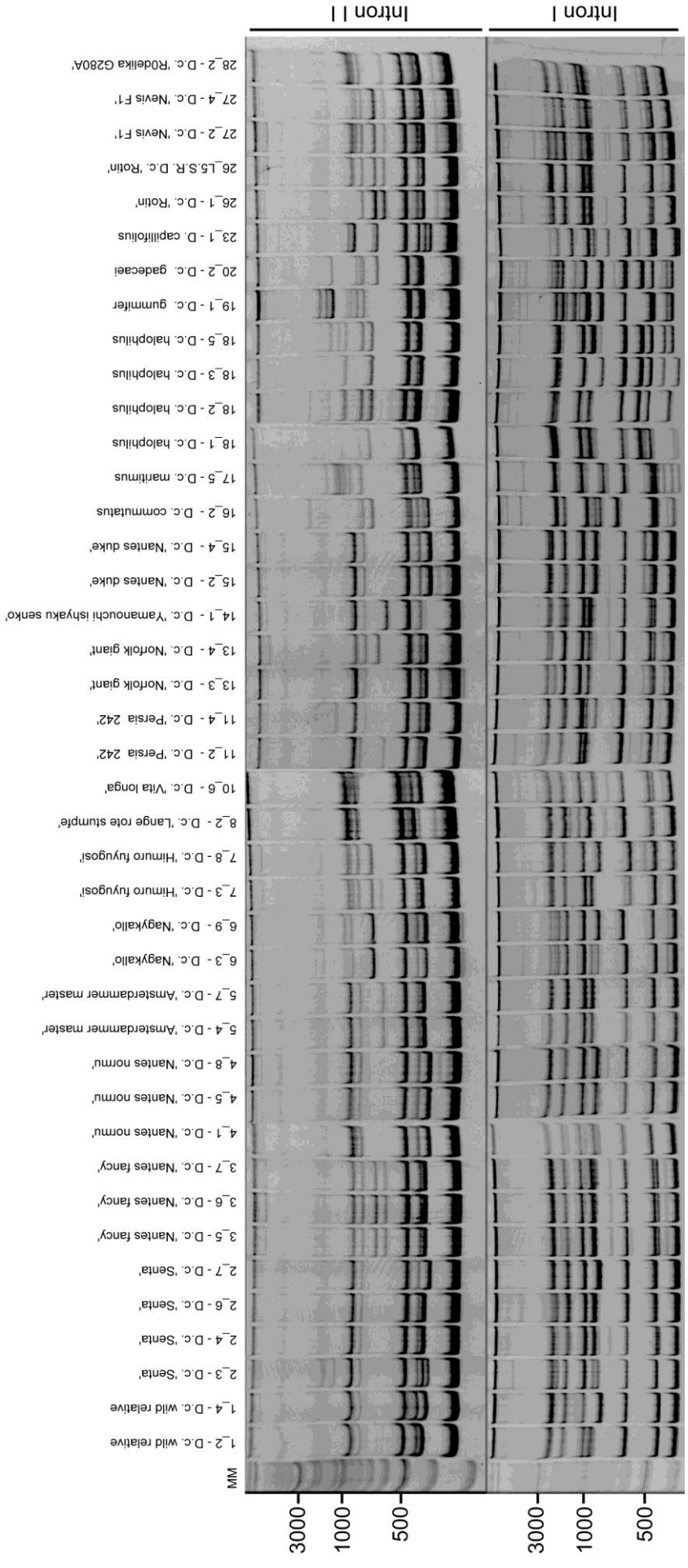


Figure 3.3 - cTBP ILP patterns characterizing the 41 cell lines evaluated (see Appendix 1). **MM** - Molecular marker - GeneRuler™ DNA Ladder Mix (100 - 10000 bp).

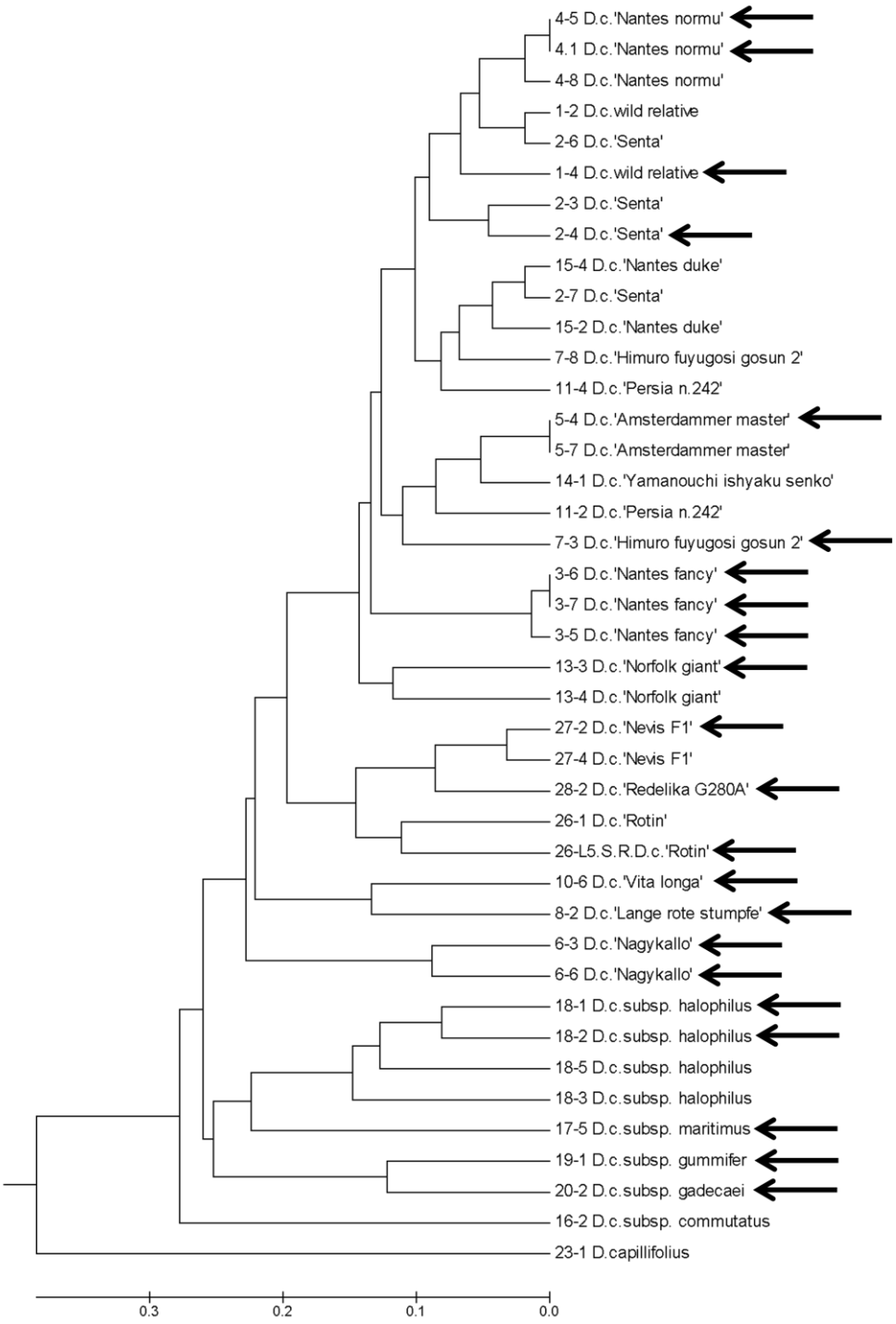


Figure 3.4 - Dendrogram representing the genetic distance of the 41 cell lines selected at the end of the first phenotyping set (see Appendix 1) assessed using the cTBP method (see section 2.2.3 and Appendices 4 and 5). The dendrogram was obtained in the FreeTree software (Hampl *et al.*, 2001) and is drawn to scale, using the UPGMA method (Sokal and Michener, 1958; Murtagh, 1984). Bootstrapping (Efron, 1979) was performed with 1000 replicates. **Arrows** - Cell lines selected for the second phenotyping set; ***D. c.*** - *Daucus carota*; **subsp.** - Subspecies.

The cell line derived from *D. capillifolius* species was the most divergent one, followed by the one from the subspecies *commutatus*. Interestingly, this last subspecies does not cluster with the remaining subspecies.

The cTBP markers grouped the cultivars in different subclusters, which may be a helpful tool to understand their breeding development. The most visible example of this, were the cell lines derived from the cultivar 'Senta', which clustered with the ones from the accessions 'Wild relative', 'Nantes duke', 'Himuro fuyugosi gosun' and 'Persia'. Other cases were also found in derived cell lines from 'Amsterdammer master', 'Nevis', 'Vita longa' and 'Yamanouchi ishyaku Senko'.

3.2.3 - Second phenotyping set for SE efficiency

The stability of the SE efficiency phenotype over time and growth phase, was evaluated on a second phenotyping set using 22 cell lines. Cell lines were selected from the ones better scored as stable on the first phenotyping set and considering cTBP genetic diversity evaluation assessment. The second set covered 17 accessions, corresponding to 68 % of the 25 initially used. On this set, 88 phenotyping experiments were assessed at four observation time points (T15, T30, T45 and T60) as in the first set, resulting in 352 phenotyping observations performed and documented (see Figure 3.5 and Appendix 1). In the second phenotyping set, cells were used during the exponential growth phase (6^{th} - 8^{th} day after subculture) in all replicates, as described in Frederico *et al*. (2009).

The cell lines 5_4, 6_3, 10_6, 18_3, 27_2 and 28_2 (five *D. c*. cultivars and one *D. c.* subspecies) considered as minimally stable on the first phenotyping set were selected for re-evaluation. The majority was classified as very efficient/efficient at the end of the first set.

As *NE*, six cell lines were selected from the first set to re-evaluation, namely: 2_4, 3_7, 4_5, 7_3, 18_1 and 19_1 (four *D. c*. cultivars and two *D. c.* subspecies). Yet, at the end of the second set, eight (2_4, 3_7, 4_5, 17_5, 18_1, 19_1 and 28_2) were found to display the phenotype in 25 phenotyping experiments. Five expressed it in a stable way. For the remaining three the phenotype proved to be unstable. The five cell lines (2_4, 3_7, 4_5, 18_1 and 19_1) phenotyped as *NE* in the second phenotyping set, had also been in the first. Cell lines 7_3, 17_5, and 28_2 showed a divergent efficiency in the second set.

As *VI*, just the cell line 6_6 (*D. c.* cultivar) was selected from the first set for re-evaluation. Yet, at the end of the second phenotyping set, four (3_5, 6_6, 7_3 and

17_5) were found to display the phenotype in 11 phenotyping experiments. Two were found to express the phenotype in a stable way. For the remaining three, the expression of the phenotype was unstable. From the two (6_6 and 3_5) phenotyped as *VI* in the second set, just 6_6 had also been classified as *VI* in the first. The cell line 17_5 showed two divergent phenotypes from *VI* in the second set and cell line 7_3 had been phenotyped as *VI* just in one out of four replicates. Both cell lines with divergent phenotypes in the second set had other phenotypes in the first.

As *I*, two cell lines were selected from the first set for re-evaluation, namely: 3_5 and 3_6 (*D. c.* cultivars). However, at the end of the second phenotyping set, two (17_5 and 18_3) cell lines different from the selected ones were found to express it on two phenotyping experiments. Both were found to express the phenotype in an unstable manner. From the ones (3_5 and 3_6) phenotyped as *I* in the first set, none were able to maintain identical phenotype in the second, in any of the four replicates performed.

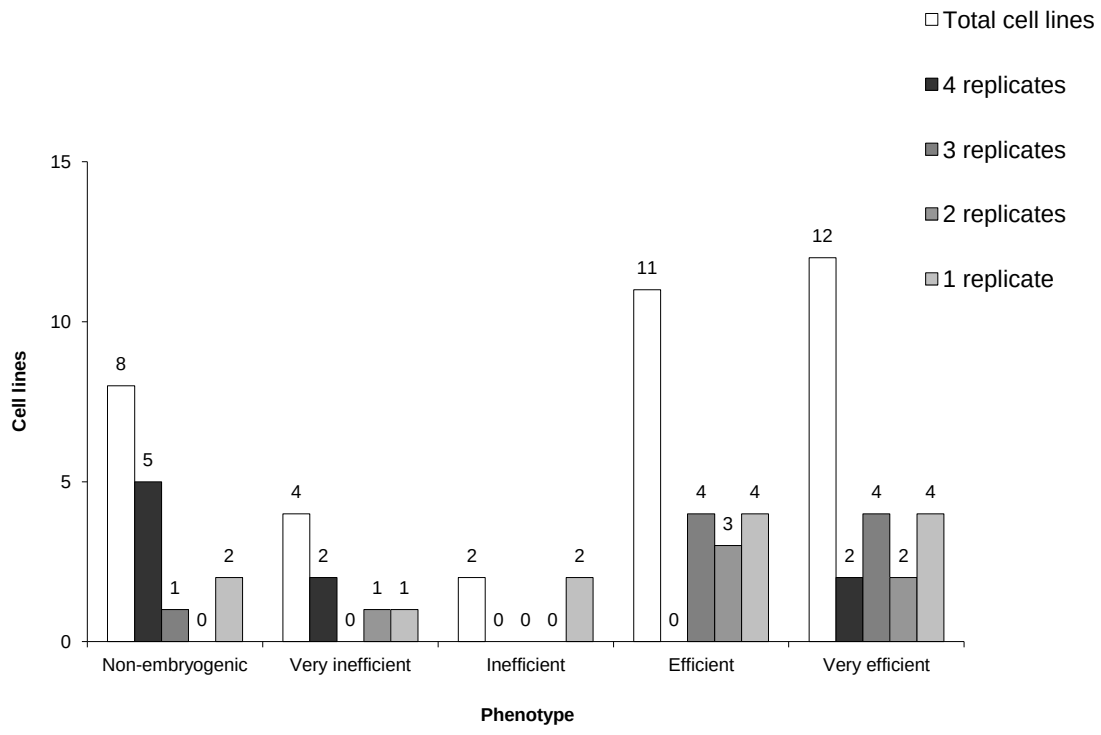


Figure 3.5 - Graphical representation of cell lines SE efficiency phenotypes obtained from the second phenotyping experiment set. The experiment (four replicates) was carried out on gelled medium B_5^- during a 60 day period and evaluation and documentation were performed at four different time points (T15, T30, T45 and T60) after SE induction. The original data used for generating the graph are provided in Appendix 1. **Grey bars** - Total number of cell lines showing an embryogenic phenotype in four (dark grey), three (mid-dark-grey), two (mid-light-grey) and one (light grey) replicates; **White bars** - Total number of cell lines detected per embryogenic phenotype, independently of the number of replicates.

As *E*, two cell lines were selected from the first set for re-evaluation, namely: 4_1 and 13_3 (*D. c.* cultivars). However, at the end of the second phenotyping set, 11 (1_4, 3_6, 4_1, 5_4, 6_3, 10_6, 13_3, 18_3, 26_L5.S.R. 27_2 and 28_2) were found to express it in 22 phenotyping experiments. All cell lines expressed the *E* phenotype in an unstable way. The two (4_1 and 13_3) cell lines phenotyped in the first set as *E*, in the second set maintained identical phenotype.

As *VE*, five cell lines were selected from the first set for re-evaluation, namely: 1_4, 8_2, 17_5, 20_2 and 26_L5.S.R. (three *D. c.* cultivars and two *D. c.* subspecies). Yet, at the end of the second phenotyping set, 12 (1_4, 3_6, 4_1, 5_4, 6_3, 8_2, 10_6, 13_3, 20_2, 26_L5.S.R., 27_2 and 28_2) were found to display the VE phenotype in 28 phenotyping experiments. Two (8_2 and 20_2) were found to express the phenotype in a stable manner on eight phenotyping replicates. For the remaining ten cell lines (1_4, 3_6, 4_1, 5_4, 6_3, 10_6, 13_3, 26_L5SR, 27_2 and 28_2) the expression of the phenotype was unstable. From the five cell lines phenotyped as *VE* in the first set, just two (8_2 and 20_2) were able to maintain identical phenotype in the second set.

3.2.4 - Cell lines selection for extreme embryogenic behaviors

At the end of the second phenotyping set, from the cell lines with the SE efficiency identical to the one observed in the first, eight were identified and selected as the most stable for extreme embryogenic phenotypes (Table 3.1). From those, four were embryogenic (8_2, 20_2, 26_L5.S.R and 27_2) (Figure 3.6) and four were non-embryogenic (2_4, 4_5, 18_1 and 19_1) (Figure 3.7).

The cell lines held as non-embryogenic expressed the same phenotype during eight phenotyping experiments, being considered as extremely stable.

From the cell lines selected as embryogenic, two (8_2 and 20_2) expressed equal phenotype during the eight phenotyping experiments performed, being classified as *VE* and qualitatively coded as +++ (high amount and quality of embryonic structures).

The cell line 27_2, showed six times the *VE* phenotype and two times the *E* phenotype, being classified as *VE*/*E* and qualitatively coded as -/+ (medium amount and quality of embryonic structures).

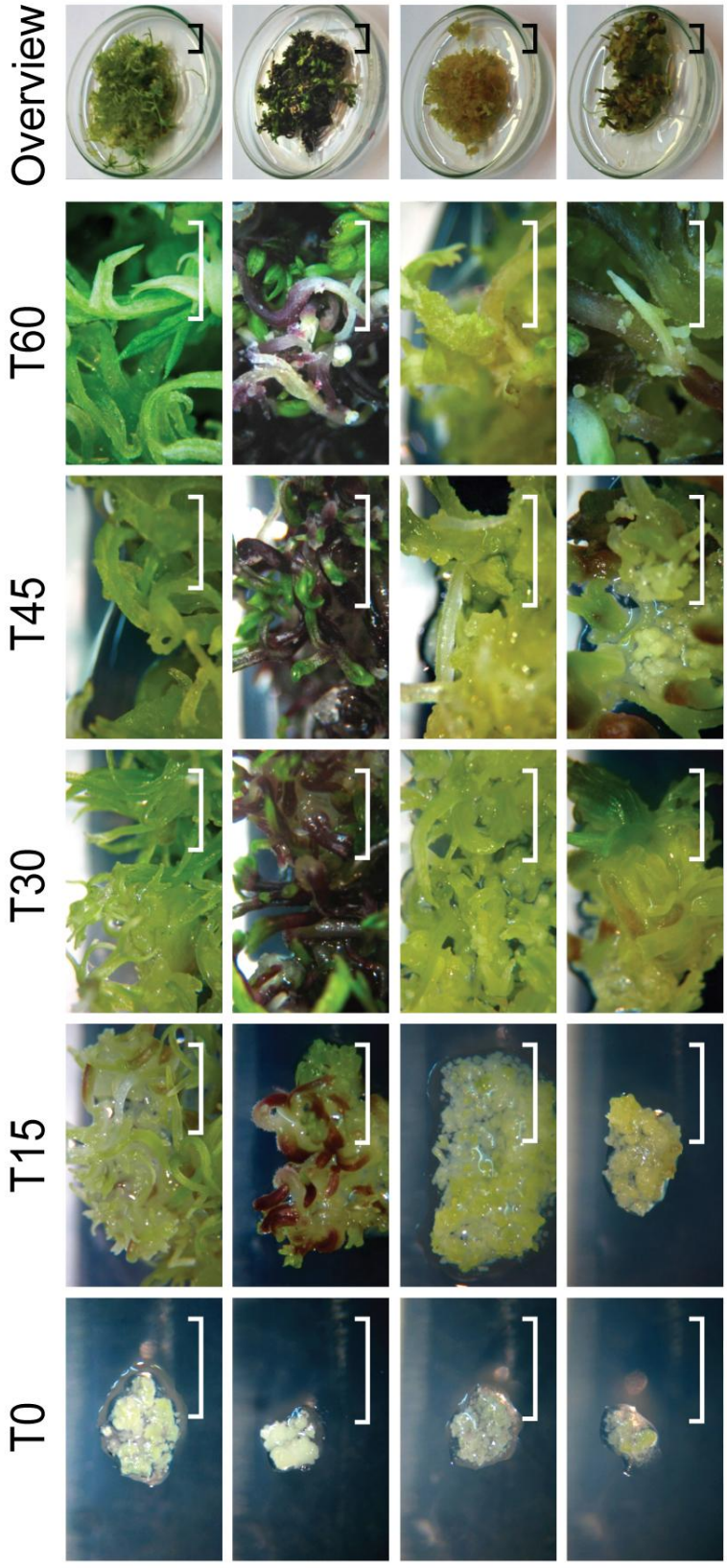


Figure 3.6 - Cell lines with highly efficient and stable embryogenic phenotype, selected at the end of both SE phenotyping experiments sets of four replicates each. Selection was performed using data presented in Appendix 1 and considering the qualitative scale presented in Appendix 2. For each cell line selected, representative images were used to illustrate the number embryonic structures at the four observation points (T15, T30, T45 and T60). Representative images were also used to illustrate inoculation (T0) and the overview at the end of the experiment. **Cell line** - Very efficient (*VE*): 8_2 - *D. c.* 'Lange rote stumpfe 1' and 20_2 - *D. c. gadecaei*; Very efficient (*VE*) / Efficient (*E*): 26_L5.S.R. - *D. c.* 'Rotin' and 27_2 - *D. c.* 'Nevis F1'; **Bar** - T0, T15, T30, T45 and T60 - 4 mm and Overview - 1 cm.

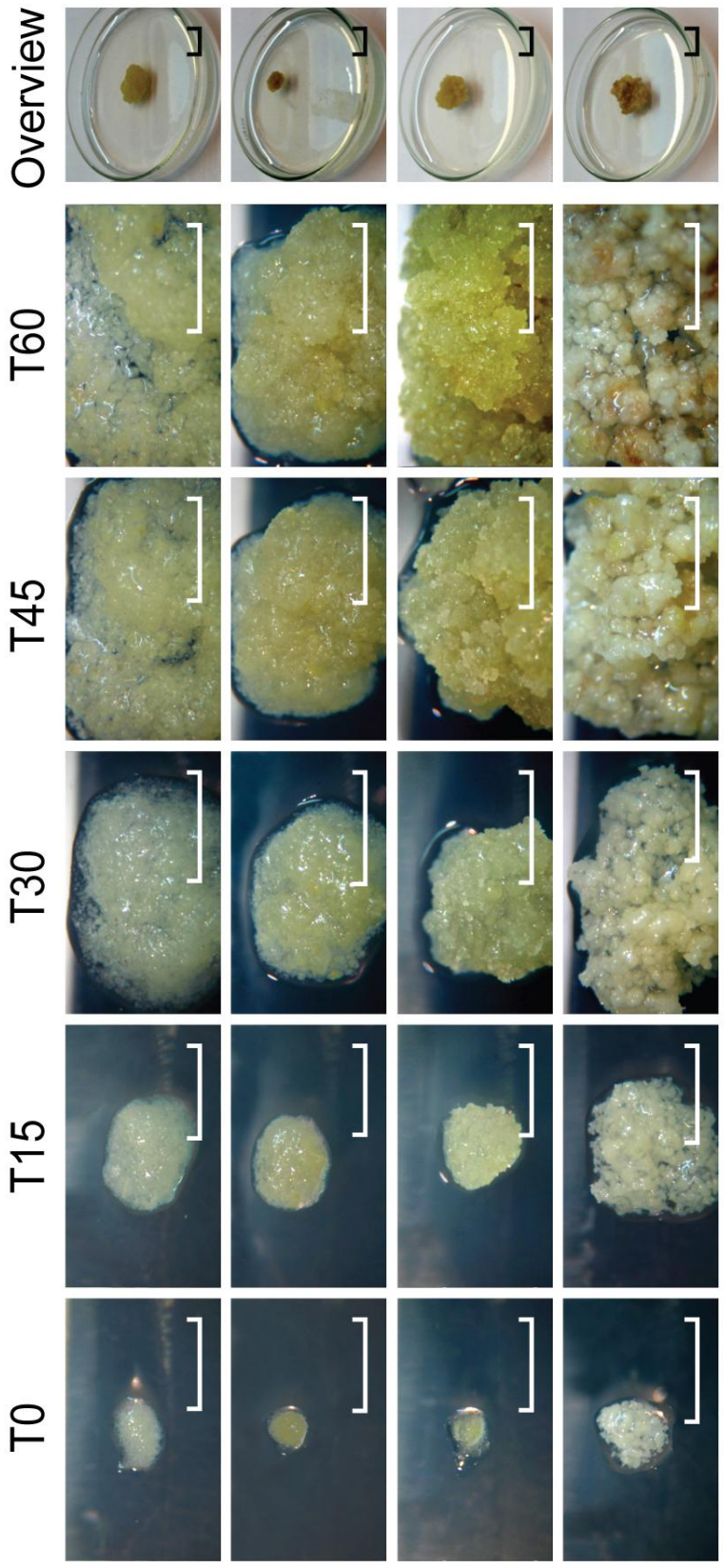


Figure 3.7 - Cell lines with highly stable non-embryogenic phenotype, selected at the end of both SE phenotyping sets of four replicates each. Selection was performed using data presented in Appendix 1 and considering the qualitative scale presented in Appendix 2. For each cell line selected, representative images were used to illustrate embryonic structures at the four observation points (T15, T30, T45 and T60). Representative images were also used to illustrate inoculation (T0) and the overview at the end of the experiment. **Cell line** - 2_4 - *D. c.* 'Senta', 4_5 - *D. c.* 'Nantes normu', 18_1 - *D. c. halophilus*, 19_1 - *D. c. gummifer*; **Bar** - T0, T15, T30, T45 and T60 - 4 mm and Overview - 1 cm.

The cell line 26_L5.S.R expressed seven times the *VE* phenotype and once the phenotype *E*, being classified as *VE/E* and qualitatively coded as -/+ (medium amount and quality of embryonic structures). This cell line was in use already in the laboratory and several preliminary experiments had already been performed with it (Frederico *et al.*, 2009a), justifying its inclusion in the present study as the reference cell line. These eight cell lines represent the final collection being used for the search of *AOX* polymorphic sites and their possible correlation with the SE efficiency phenotype, and were kept in triplicate, as stock material, by periodic subculture at the Plant Breeding and Biotechnology Laboratory of the University of Évora.

3.2.5 - Cell lines ploidy determination

Due to *in vitro* stressful conditions, knowing the mutagenic effect of the synthetic auxin (2,4-D) in the cells and aiming to perform a deep characterization of the eight cell lines selected, the ploidy level and the relative DNA amount was assessed by flow cytometry. Table 3.2 synthesizes the flow cytometry results, determined from collected peaks data from each control leaf (2n) and related cell line. The corresponding full data are provided in Appendix 6.

Table 3.2 - Results from flow cytometry analyses performed using the eight selected cell lines considering both phenotyping sets. The presented results concerns the relative DNA amount determined for each control leaf (2n) and related cell line, cell line ploidy level and percentage of cells present in each line with double amount of DNA (cells in the G2/M phase transition). Original peak data used for the generation of this table is provided in Appendix 6. **Cell line** - 2_4 - *D. c*. 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

	Relative DNA ratio (Internal standard – *B. sempervirens*)			
Cell line	**Leaf control** (2n=2x=18)	**Cell line**	**Ploidy**	**G2 cells** (%)
2_4	0,76	0,83	2X	36
4_5	0,74	1,46	3,5X	42
8_2	0,75	2,72	6,5X	5
18_1	0,91	0,93	2X	30
19_1	0,76	0,82	2X	27
20_2	0,78	0,83	2X	16
26_L5.S.R.	0,74	2,67	6,5X	12
27_2	0,75	0,79	2X	19

From the eight leaves used as control, seven presented comparable relative DNA amounts when compared with the internal standard (*B. sempervirens*). The control

leaf belonging to accession 18 (*D. c. halophilus*) was the only one diverging in the relative DNA amount in comparison with the internal standard.

Five (2_4, 18_1, 19_1, 20_2 and 27_2) of the cell lines under evaluation showed a relative DNA amount comparable with the leaf control, all of them being diploid. The remaining three cell lines were polyploids, with 8_2 and 26_L5.S.R showing a relative DNA amount 6,5X higher than the control leaves and cell line 4_5 showing a 3,5X higher. All non-embryogenic cell lines presented a high percentage of cells in the G2/M transition phase, when compared with the embryogenic ones.

3.3 - *AOX* genes in the selected cell lines

Sequence diversity of the eight selected cell lines was assessed through genomic *AOX* amplification (from start to end codon) using PCR, in order to evaluate the potential role of *AOX* genes polymorphic sites on SE efficiency phenotypes. For the amplification of *AOX* genes specific primers were used. Amplified *AOX* variants were confirmed by resequencing and evaluated using bioinformatic tools as described in sections 2.2.5 and 2.2.6. The identity of the amplified *AOX* sequences had been assessed at the NCBI database using the BLASTx and BLASTn algorithms. The similarity with previously published *Daucus AOX* sequences ranged from 95 to 100 % at the nucleotide level and from 94 to 100 % at the protein level. The *AOX2b* sequences similarity percentage only considers the exon sequences, because no intron sequences had previously been issued for this gene in the NCBI database. Details on the structural analysis performed on the identified AOX sequences are available in Appendix 3 (A - *AOX1*, B - *AOX2a* and C - *AOX2b*) and reported in the next sections.

Table 3.3 - Cell lines selected for extreme embryogenic efficiency phenotypes (VE, VE/E and NE). **Cell line** - 2_4 - *D. c*. 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

Cell line	Embryogenic phenotype
2_4	Non-embryogenic
4_5	Non-embryogenic
8_2	Embryogenic - VE
18_1	Non-embryogenic
19_1	Non-embryogenic
20_2	Embryogenic - VE
26_L5.S.R	Embryogenic - VE/E
27_2	Embryogenic - VE/E

3.3.1 - Amplified *AOX1* characterization

A total of 77 sequences were successfully amplified from the eight cell lines evaluated, representing 11 different polymorphic sequences (Appendix 3A). Cell lines 2_4 and 27_2 were found to have two and three different *AOX1* sequences, respectively. The remaining six cell lines are represented by a unique *AOX1* sequence type. The 11 *AOX1* polymorphic sequences were submitted to the NCBI database, named as: Senta_AOX1_51 (KX664821), Senta_AOX1_50 (KX664822), Nantes_normu_AOX1_1 (KX664823), L_r_stumpfe_AOX1_4 (KX664824), halophilus_AOX1_1 (KX664825), gummifer_AOX1_4 (KX664826), gadecaei_AOX1_1 (KX664827), Rotin_AOX1_1 (KX664828), Nevis_AOX1_1 (KX664829), Nevis_AOX1_47 (KX664830) and Nevis_AOX1_52 (KX664831). *AOX1* sequence length varies from 1789 bp in cell line 18_1 to 1865 bp in cell lines 2_4, 4_5 and 26_L5.S.R. Structurally, sequences have three exons and two introns. Exon 1 was found to be variable in contrast to exons 2 and 3, which were stable across cell lines. Both introns were variable in size. ORFs were also variable in size, resulting in 3 amino acid sequence size variants.

The clustering analysis performed using the 11 *AOX1* sequences, allowed the identification of four *AOX1* clusters (Figure 3.8), being two of them represented by a single sequence (*D. c. halophilus* and *D. c.* 'Senta' 50).

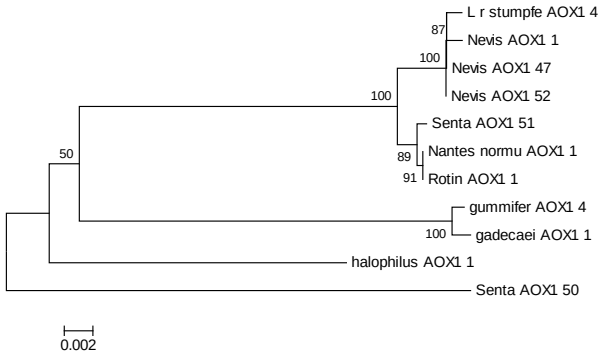


Figure 3.8 - Clustering analyses of the 11 *AOX1* nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 1949 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c*. 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

The other 'Senta' sequence identified (coded with the number 51), clusters with the remaining *D. c.* cultivars sequences ('Nantes normu', 'Lange rote stumpfe', 'Rotin' and 'Nevis F1'). The subspecies *gummifer* and *gadecaei* sequences, both with a unique sequence, form the remaining group.

3.3.2 - Amplified *AOX2a* characterization

A total of 103 sequences were successfully amplified from the eight cell lines evaluated, representing 24 different polymorphic sequences (Appendix 3B).

For the sequences 'Senta' 56 and gadecaei 7, the nucleotide sequence corresponding to the beginning of exon 1 was missing, although the remaining structure was amplified (see Appendix 3B). Both sequences were excluded from the subsequent studies concerning polymorphism exploitation, being considered for further evaluations only the remaining 22. The 22 *AOX2a* polymorphic sequences were submitted to the NCBI database, named as: Senta_AOX2a_47 (KX664832), Nantes_normu_AOX2a_8 (KX664833), L_r_stumpfe_AOX2a_3 (KX664834), L_r_stumpfe_AOX2a_17 (KX664835), halophilus_AOX2a_L56 (KX664836), halophilus_AOX2a_L146 (KX664837), halophilus_AOX2a_L177 (KX664838), halophilus_AOX2a_L214 (KX664839), halophilus_AOX2a_L219 (KX664840), halophilus_AOX2a_S1 (KX664841), gummifer_AOX2a_4 (KX664842), gummifer_AOX2a_14 (KX664843), gummifer_AOX2a_64 (KX664844), gadecaei_AOX2a_3 (KX664845), gadecaei_AOX2a_31 (KX664846), gadecaei_AOX2a_35 (KX664847), Rotin_AOX2a_L13 (KX664850), Rotin_AOX2a_L101 (KX664851), Rotin_AOX2a_L128 (KX664848), Rotin_AOX2a_L142 (KX664849), Rotin_AOX2a_S5 (KX664852) and Nevis_AOX2a_27 (KX664853).

Cell lines 2_4, 4_5 and 27_2 were represented each one by a unique *AOX2a* sequence variant. Cell lines 8_2, 18_1, 19_1, 20_2 and 26_L5.S.R were represented by a different number of variant sequences each, namely: 2, 6, 3, 3 and 5, respectively. Sequence length varied from 4911 bp in cell line 20_2 to 5315 bp in cell line 26_L5.S.R. Structurally, sequences have four exons and three introns. Exon 1 was found to be three nucleotides shorter in cell lines 19_1 and 20_2, both *D. c.* subspecies, contrarily to exons 2, 3 and 4 which were stable across cell lines. Introns

1, 2 and 3 were highly variable in size. ORFs were also shorter in the cell lines with shorter exon 1, resulting in two amino acid sequence size variants.

After performing a clustering analysis using the 22 *AOX2a* sequences, three major clusters were identified, as represented in Figure 3.9. Sequences belonging to *D. c gadacaei* and *D. c. gummifer* form a consistent cluster apart from all the other cell lines *AOX2a* sequences. *D. c. halophilus* sequences, including the long (L) and short (S) types, all group together, but two 'Rotin' L type sequences also belong to this cluster. The third cluster joins together the remaining sequences belonging to the *D. c.* cultivars ('Senta', Nantes normu', Lange rote stumpfe', 'Nevis F1' and 'Rotin'), including another L type sequence from 'Rotin', as well as the S type.

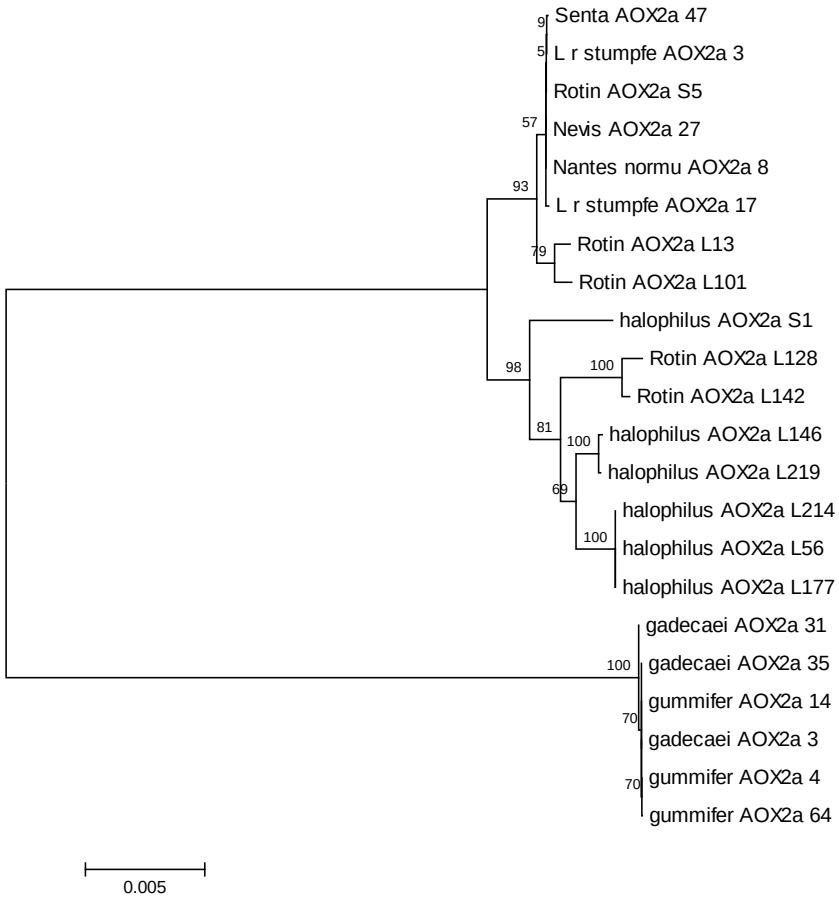


Figure 3.9 - Clustering analyses of the 22 *AOX2a* nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 5385 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

3.3.3 - Amplified *AOX2b* characterization

A total of 110 sequences were successfully amplified from the eight cell lines evaluated, representing 14 different polymorphic sequences (Appendix 3C). The 14 *AOX2b* polymorphic sequences were submitted to the NCBI database, named as: Senta_AOX2b_57 (KX664854), Nantes_normu_AOX2b_1_16 (KX664855), Nantes_normu_AOX2b_3_21 (KX664856), L_r_stumpfe_AOX2b_0_32 (KX664857), L_r_stumpfe_AOX2b_1_34 (KX664858), L_r_stumpfe_AOX2b_1_45 (KX664859), halophilus_AOX2b_17 (KX664860), gummifer_AOX2b_26 (KX664861), gummifer_AOX2b_31 (KX664862), gadecaei_AOX2b_5 (KX664863), Rotin_AOX2b_17 (KX664864), Nevis_AOX2b_47 (KX664865), Nevis_AOX2b_56 (KX664866) and Nevis_AOX2b_57 (KX664867).
Cell lines 4_5, 8_2, 19_1 and 27_2 were found to have a different number of *AOX2b* sequence variants each, namely: 2, 3, 2 and 3, respectively. The remaining four cell lines are represented by a unique *AOX2b* sequence. Sequence length varied from 1885 bp in cell lines 8_2, 18_1 and 27_2 to 2344 bp in cell line 2_4. Structurally, sequences have four exons and three introns. Exons were stable across cell lines. Intron 1 was variable in size, but the other two were stable, with the exception of intron 3 from cell line 20_2 (*D. c. gadecaei*), which was larger. ORFs were stable in size, resulting in a single amino acid sequence size variant.

The clustering analysis performed using the 14 *AOX2b* sequences identified, allowed the identification of five clusters of sequences (Figure 3.10), one of them being represented by the single sequence identified on that cell line (20_2 - *D. c. gadecaei*).
All remaining clusters contained sequences belonging to more than one accession cell line, showing a high degree of *AOX2b* sequence diversity among the accession cell lines under study. Three *AOX2b* sequences were identified from *D. c.* 'Nevis F1', each one grouping within a different cluster, and therefore being present in three different sequence clusters. For *D. c.* 'Lange rote stumpfe', also three sequences were identified, two of them grouping in a cluster and the remaining one on another. Both *D. c.* 'Nantes normu' and *D. c. gummifer*, contained two *AOX2b* sequences, each of them, grouping into a different cluster. The *AOX2b* sequences identified from

subspecies all grouped together with those identified from cultivars, in contrast with what happen in the case of *AOX1* and *AOX2a* sequences.

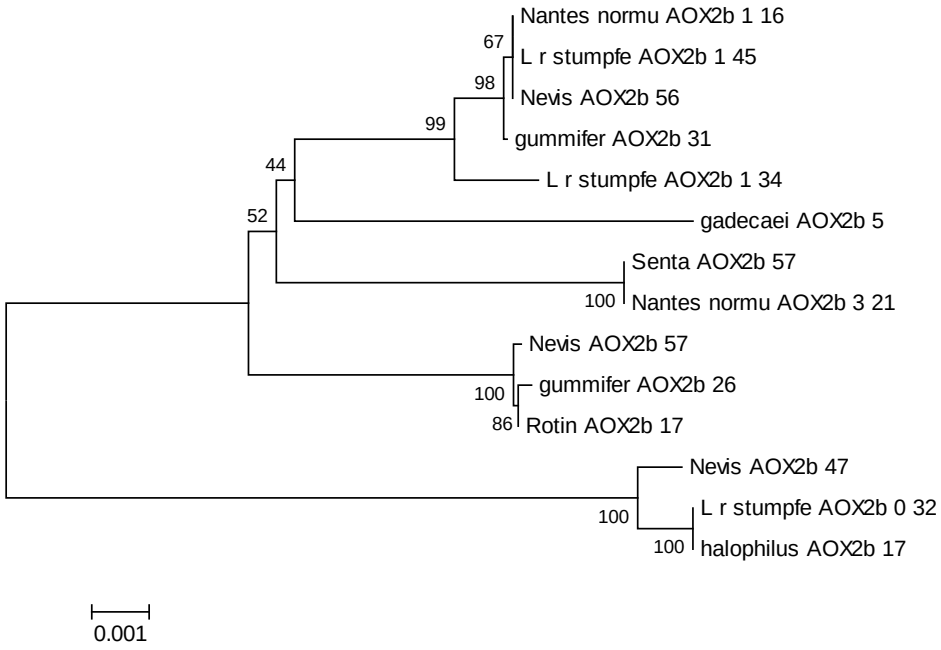


Figure 3.10 - Clustering analyses of the 14 *AOX2b* nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analysis distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 2423 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

3.4 - Evaluation of *AOX* polymorphic sites

The clustering analysis described above, using the complete sequences of the *AOX* genes under investigation, did not permit the recognition of a discrete group of sequences containing a specific polymorphic site associated with the SE efficiency phenotypes. Therefore, clustering analyses were next performed using specific sequence regions (coding and non-coding) and the deduced amino acid sequences, aiming to better infer the potential role of the polymorphic sites identified in the alignments in relation with the SE efficiency phenotypes. In the same way a search for (post)-transcriptional regulatory elements was performed in specific sites with a high degree of polymorphic positions.

3.4.1 - Analysis of *AOX* coding region

The clustering analysis, performed using the *AOX1*, *2a* and *2b* ORFs and corresponding amino acid sequences, is presented in Figures 3.11 and 3.12.
The analysis showed that ORFs and amino acid sequences belonging to each gene evaluated, clustered together with the sequences belonging to the same gene. Three clusters were detected in each analysis (Figures 3.11 and 3.12), each one grouping the sequences belonging to the specific gene under evaluation.
AOX1 sequences (ORFs and amino acids) form two main clusters on each analysis (Figures 3.11 and 3.12). One of them contains only sequences from *D. c.* cultivars (*D. c.* 'Nevis F1' and *D. c.* 'Lange rote stumpfe'), whereas the other one contains sequences from *D. c.* subspecies (*D. c. halophilus*, *D. c. gummifer* and *D. c. gadecaei*) and *D. c.* cultivars (*D. c.* 'Senta', *D. c.* 'Nantes normu' and *D. c.* 'Rotin'). When ORFs were analyzed, four *D. c.* cultivars sequences (*D. c.* 'Senta' 50 and 51, *D. c.* 'Nantes normu' 1 and *D. c.* 'Rotin' 1) clustered with the ones from *D. c.* subspecies, however, when amino acid sequences were used, only two (*D. c.* 'Senta' 51 and *D. c.* 'Rotin' 1) of them remained on that cluster.

The clustering analysis of *AOX2a* sequences (ORFs and amino acids) revealed the existence of several clusters (Figure 3.11 and 3.12). A clear separation amongst the *D. c.* cultivar sequence cluster and the two clusters grouping subspecies sequences was identified, when ORFs were used. The cluster grouping *D. c. halophilus* ORF sequences contained two subclusters, indicating the presence of two sequence variants, one with the S and L type (S1, L146 and L219) and the other only with L type (L56, L177, L214) sequences. The *AOX2a* ORF sequences from *D. c. gummifer* and *D. c. gadecaei* formed a unique cluster, and the same was true for, the sequences belonging to *D. c*. cultivars. When AOX2a amino acid sequences were used for clustering analysis, *D. c. halophilus* sequences remained in a separated cluster with two subclusters. The *gummifer* / *gadecaei* cluster, received a member from the *D. c.* cultivars (*D. c.* 'Senta' 56). The *D. c.* cultivars amino acid sequences form the remaining cluster, from which two *D. c.* 'Rotin' L type sequences (L13 and L128) appeared separated.

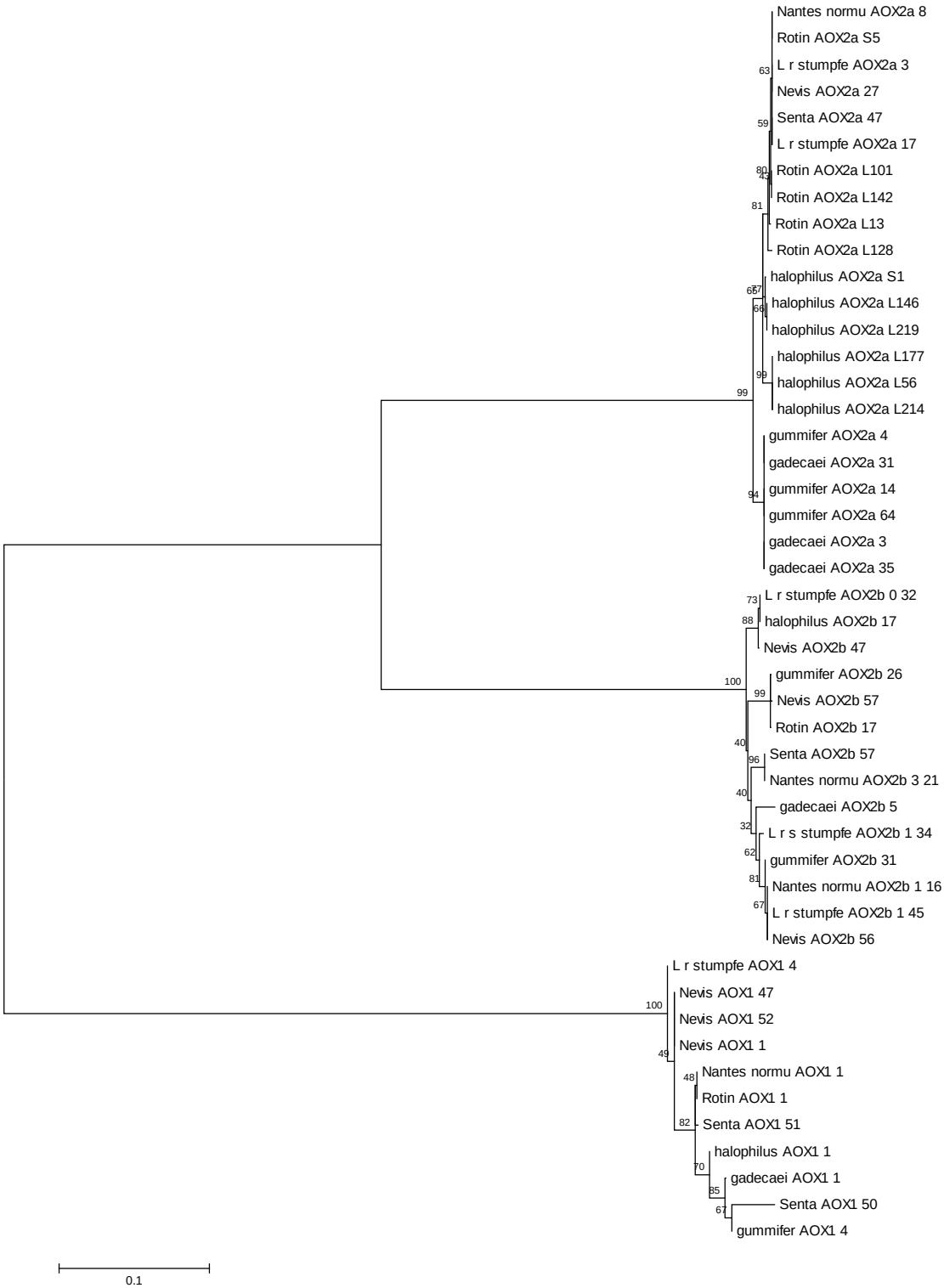


Figure 3.11 - Clustering analyses of 47 *AOX1*, *2a* and *2b* open reading frames nucleotide sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein,1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 1020 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

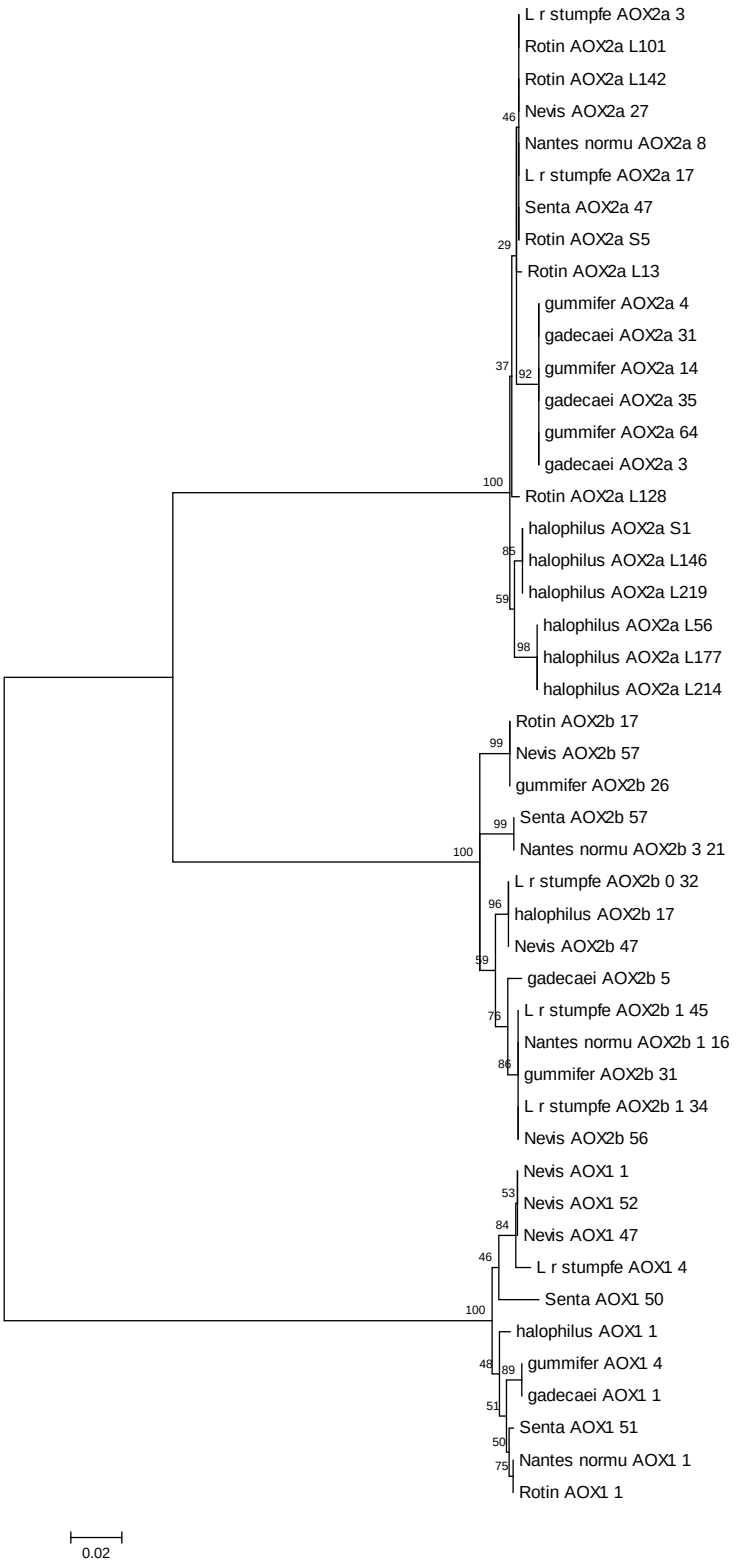


Figure 3.12 - Clustering analyses of 47 AOX1, 2a and 2b amino acid sequences identified in the 8 selected cell lines with extreme embryogenic behaviors. The clustering analyses were inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 346 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

In the case of *AOX2b* sequences (ORFs and amino acids), three clusters were identified (Figures 3.11 and 3.12). Sequences from *D. c.* cultivars and subspecies clustered together, with no clear separation amongst them.

Cell lines with more than one *AOX2b* ORF sequence, had representatives on more than one cluster (*D. c.* 'Nevis F1', *D. c.* 'Lange rote stumpfe' and *D. c. gummifer*). In the same way, when amino acid sequences were used, cell lines with more than one sequence had representatives on more than one cluster (*D. c.* 'Nevis F1', 'Nantes normu' and *D. c. gummifer*). The fact that all *D. c.* 'Lange rote stumpfe' amino acid AOX2b sequences clustered together, indicates the presence of synonymous mutations in the *AOX2b* ORF sequences isolated from that cell line.

The analyses performed in the coding region of the *AOX* genes revealed the presence of several mutations in all exons. For identification of mutated positions and the study of its effect on the translated amino acid, ORFs of each *AOX* gene were aligned and the polymorphic positions identified. The result of the mutations detected on each gene and the effect on the translated amino acid are presented in the Appendix 7 (7A - *AOX1*, 7B - *AOX2a* and 7C - *AOX2b*) and 8.

Globally, the mutations identified were SNPs (non-synonymous SNPs - nsSNPs and synonymous SNPs - sSNPs) and InDels. The overall number and type of mutations identified for each gene is presented in Table 3.4.

The *AOX1* gene was the one with the highest level of variation, with 38 SNPs and two InDels, followed by the *AOX2b* with 34 SNPs and finally the *AOX2a* with 25 SNPs and a single InDel.

Table 3.4 - Distribution of mutations (nsSNPs, sSNPs and InDels) across *AOX1*, *2a* and *2b* exons. Full data used to construct the table are provided in the Appendix 7. * - *AOX1* does not have exon 4.

			Exon 1	Exon 2	Exon 3	Exon 4	**Total**
AOX1	SNPs	Non-synonymous	12	1	1	*	**14**
		Synonymous	18	6	0	*	**24**
	InDels		2	0	0	*	**2**
AOX2a	SNPs	Non-synonymous	8	2	1	0	**11**
		Synonymous	5	2	6	1	**14**
	InDel		1	0	0	0	**1**
AOX2b	SNPs	Non-synonymous	12	3	0	0	**15**
		Synonymous	3	2	13	1	**19**

Exon 1 always showed the highest rate of mutations in all the three genes under study being InDels only found on this exon. nsSNPs were also concentrated mainly

on exon 1, but sSNPs were found more widespread across the other exons, like in the case of *AOX2b* exon 3 with 13 SNPs.
The sSNPs number was always higher than the nsSNPs one on the three *AOX* genes assessed.

Aiming to understand the putative functional role of polymorphic sites detected in the *AOX* ORFs nucleotide variants, the enrichment of regulatory element motifs was evaluated using two software, Nsite-PL and ScanWM-P (see section 2.2.7) (Table 3.5), with the. Both software were able to identify differences at the level of regulatory elements number and their relative enrichment due to the existence of polymorphic sites. Differences were identified amongst accessions and also amongst variant sequences obtained within accessions.
The Nsite-PL for *AOX1* just detected differences amongst *Daucus* subspecies (*halophilus* - 18_1, *gummifer* - 19_1 and *gadecaei* - 20_2) and the *Daucus* species sequences. ScanWM-P was able to identify a higher number of weight matrixes and motifs using the same sequences, detecting differences amongst accessions. Differences were also detected within the two variants (50 and 51) of *AOX1* from accession 'Senta' (2_4), at the level of weight matrices and at the level of motifs detected. The detected putative regulatory elements (Table 3.5) which were found in *AOX* ORF nucleotide sequences polymorphic positions, are represented in Appendix 9 (9A - *AOX1*, 9B - *AOX2a* and 9C - *AOX2b*).

From the elements initially identified in the *AOX1* ORFs, the Nsite-PL identified elements that occupy positions on seven polymorphic sites and the ones identified by the ScanWM-P were found on 29. The sequence 50 identified from cell line 2_4 (*D. c.* 'Senta') was the one with the largest number of specific regulatory elements, being two of them identified when the Nsite-PL was used and ten when the used software was the ScanWM-P. All the regulatory elements identified by the Nsite-PL were detected just once in the *AOX1* ORF sequences, but when the ScanWM-P was used, the sequences were found enriched with the W box (W) and HVA1 motif from barley GCCGAC gene (A1) elements, which were found repeatedly along the sequence at polymorphic sites. The regulatory elements identified by both software were found distributed along the sequence with no clear distinction amongst the embryogenic or non-embryogenic ones.

Table 3.5 - Enrichment of putative regulatory elements in the identified *AOX1*, *2a* and *2b* ORF nucleotide sequences from the 8 selected cell lines assessed using Nsite-PL and ScanWM-P software (see section 2.2.7). **bp** - Base pair; **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed (see Appendix 3); **Grey cells** - Sequences with different levels of identified regulatory elements/weight matrixes/motifs enrichment; **Nr.** - Number.

			Lenght (bp)	Nsite-PL (Nr.)		ScanWM-P (Nr.)	
	Cell line	Clone		Regulatory elements	Motifs	Weight Matrixes	Motifs
AOX1	**2_4**	**50**	***981***	27	28	50	88
		51	***951***	27	28	42	71
	4_5	**1**	***951***	27	28	43	72
	8_2	**4**	***942***	27	28	48	77
	18_1	**1**	***951***	26	27	44	77
	19_1	**4**	***951***	26	27	44	75
	20_2	**1**	***951***	26	27	45	76
	26_L5.S.R.	**1**	***951***	27	28	43	72
	27_2	**1**	***942***	27	28	47	75
		47	***942***	27	28	47	75
		52	***942***	27	28	47	75
AOX2a	**2_4**	**47**	***1017***	16	16	17	29
	4_5	**8**	***1017***	16	16	17	29
	8_2	**3**	***1017***	16	16	17	29
		17	***1017***	16	16	17	29
	18_1	**S1**	***1017***	14	14	15	28
		L56	***1017***	17	17	15	29
		L146	***1017***	14	14	15	28
		L177	***1017***	17	17	15	29
		L214	***1017***	17	17	15	29
		L219	***1017***	14	14	15	28
	19_1	**4**	***1014***	18	18	17	30
		14	***1014***	18	18	17	30
		64	***1014***	18	18	17	30
	20_2	**3**	***1014***	18	18	17	30
		31	***1014***	18	18	17	30
		35	***1014***	18	18	17	30
	26_L5.S.R.	**S5**	***1017***	16	16	17	29
		L13	***1017***	17	17	17	30
		L101	***1017***	16	16	17	30
		L128	***1017***	17	17	17	28
		L142	***1017***	16	16	17	30
	27_2	**27**	***1017***	16	16	17	29
AOX2b	**2_4**	**57**	***960***	18	19	22	31
	4_5	**1_16**	***960***	17	18	22	31
		3_21	***960***	18	19	22	31
	8_2	**0_32**	***960***	13	14	22	31
		1_34	***960***	18	19	22	32
		1_45	***960***	17	18	22	31
	18_1	**17**	***960***	13	14	22	31
	19_1	**2_26**	***960***	15	16	26	39
		2_31	***960***	17	18	22	31
	20_2	**5**	***960***	16	17	24	34
	26_L5.S.R.	**17**	***960***	15	16	26	39
	27_2	**47**	***960***	14	15	22	31
		56	***960***	17	18	22	31
		57	***960***	15	16	26	39

In the case of *AOX2a* sequences, both software were able to detected differences amongst accessions, as well amongst variants within accessions (*halophilus* and 'Rotin'). The number of regulatory elements detected by the two software was almost identical, but the level of enrichment was higher when ScanWM-P was used.

The Nsite-PL detected regulatory elements on six polymorphic positions and the ScanWM-P just in one. All the elements detected on polymorphic positions were detected just once. The elements found were grouped mainly by accession and no clear distinction amongst SE efficiency phenotype was evident.

As for *AOX2a*, also for *AOX2b* sequences, both software were able to detect differences amongst accessions, as well as within accessions variant sequences (‘Nantes normu’, ‘Lange rote stumpfe‘ and *gummifer*). However, the Nsite-PL identified a higher degree of variation than ScanWM-P.
On the *AOX2b* variant sequences, the Nsite-PL identified regulatory elements on seven polymorphic positions and the ScanWM-P on 14. The elements identified by the Nsite-P were single detections and no enrichment was detected. On the other hand, some of the elements identified by the ScanWM-P were detected more than once on the *AOX2b* ORF sequences. The ABA response (ABRE) and early methionine 1b (Em1b) elements were the ones found enriched in the sequences. Interestingly, the mitosis-specific activator (MSA), ABRE3 and Emb1 identified with the ScanWM-P, were found associated mainly with sequences belonging to embryogenic cell lines due to the existence of two sSNPs in the exon 3 in the positions 22 and 31 (see Appendix 7C). However, those elements were also detected associated with two sequences from the non-embryogenic accessions *halophilus* and *gummifer* (see Apppendix 9C).

After evaluated the position of the regulatory elements identified in the *AOX* ORF sequences with both software, it was possible to detect putative regulatory elements associated with the presence of polymorphic sites in all three genes. Although detected, none were able to establish a clear correlation between their occurrence and the cell line ability to perform SE. However, some of the results obtained for the *AOX2b* with ScanWM-P may be helpful for correlating polymorphic positions with the identified SE efficiency phenotype.

In order to assess if the deduced AOX protein could really function in the mitochondria, the webserver TargetP (see section 2.2.7) was used to predict its subcellular localization. According to the predictions performed, all AOX1, 2a and 2b amino acid sequences identified will be targeted to the mitochondria (Table 3.6). The mitochondrial targeting peptide (mTP) scores varied from 0,961 to 0,596, indicating the high probability of mitochondrial importing. The targeting peptide (TP) length varies from 20 to 61. The correspondent TP cleavage site pre-sequences are presented in Table 3.6 and represented graphically in Appendix 8.

Table 3.6 - Subcellular localization of the identified AOX amino acid sequences according to the results obtained on the TargetP 1.1 Server (see section 2.2.7). **aa** - Amino acid; **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed (see Appendix 3); **mTP** - Mitochondrial targeting peptide score; **TP** - Targeting peptide pre-sequence length.

Gene	Cell line	Clone	Peptide (aa)	mTP	TP length (aa)	Targeting peptide sequence
AOX1	2_4	50	326	0,89	41	MMMTRGTSRVARLTTADRLFSAVKGAAAESEKFPVMGVRWR
		51	316	0,949		MMMTRGTSRVARLTMGGRLFSAVKGAAAEGEKFPVMGVRWR
	4_5	1	316	0,952	41	MMMTRGTSRVARLTMGGRLFSAVKGAAAESEKFPVMGVRWR
	8_2	4	313	0,961	20	MMMTRGTSRVARFTTAGRLF
	18_1	1	316	0,948	41	MMMTRGTSRVARLTMGGRLFSAVKGAAAEGEKFPVMGVRWR
	19_1	4	316	0,949	41	MMMTRGTSRVARLTMGGRLFSAVKGAAAEGEKFPVMGVRWR
	20_2	1	316	0,949	41	MMMTRGTSRVARLTMGGRLFSAVKGAAAEGEKFPVMGVRWR
	26_L5.S.R	1	316	0,952	41	MMMTRGTSRVARLTMGGRLFSAVKGAAAESEKFPVMGVRWR
	27_2	1	313	0,96	20	MMMTRGTSRVARFTTAGRLF
		47				MMMTRGTSRVARFTTAGRLF
		52				MMMTRGTSRVARFTTAGRLF
AOX2a	2_4	47	338	0,791	61	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
	4_5	8	338	0,791	61	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
	8_2	3	338	0,791	61	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
		17		0,791		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
	18_1	S1	338	0,758	61	MNHLLAKSVMRRLISGGSSIRSASPAPSLTIFRAVTDSATMRRESLVYVRGGGVELMKRMM
		L56		0,814		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRVVTDSATMRRESLVYVRGGGVELMKRMM
		L146		0,758		MNHLLAKSVMRRLISGGSSIRSASPAPSLTIFRAVTDSATMRRESLVYVRGGGVELMKRMM
		L177		0,814		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRVVTDSATMRRESLVYVRGGGVELMKRMM
		L214		0,814		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRVVTDSATMRRESLVYVRGGGVELMKRMM
		L219		0,758		MNHLLAKSVMRRLISGGSSIRSASPAPSLTIFRAVTDSATMRRESLVYVRGGGVELMKRMM
	19_1	4	337	0,803	43	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
		14				MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
		64				MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
	20_2	3	337	0,803	43	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
		31				MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
		35				MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATVRR
	26_L5.S.R	S5	338	0,791	61	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
		L13		0,791		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
		L101		0,791		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
		L128		0,835		MNHLLAKSVMRRLISGGGSIRSASPAPALTIFRAVTESATARRESLVYVRGGGVELMKRMM
		L142		0,791		MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
	27_2	27	338	0,791	61	MNHLLAKSVMRRLISGGGSIRSASPAPSLTIFRAVTESATARRESLVYVRGGGVELMKRMM
AOX2b	2_4	57	319	0,633	20	MNQVVARSVIRRLINSQKSP
	4_5	1_16	319	0,596	20	MNQMVARSVIRRLINSQKSP
		3_21		0,633		MNQVVARSVIRRLINSQKSP
	8_2	0_32	319	0,605	20	MNQMVARSVIRRLINSQKSP
		1_34		0,596		MNQMVARSVIRRLINSQKSP
		1_45		0,596		MNQMVARSVIRRLINSQKSP
	18_1	17	319	0,605	20	MNQMVARSVIRRLINSQKSP
	19_1	26	319	0,597	20	MNQVVARSVIRRLINSQKSP
		31		0,596		MNQMVARSVIRRLINSQKSP
	20_2	5	319	0,724	55	MNQMVARSVIRRLINRQKSPISTFRSHDDIAIANRQRLGIIGGGARVFGTRMMSA
	26_L5.S.R	17	319	0,597	20	MNQVVARSVIRRLINSQKSP
	27_2	47	319	0,605	20	MNQMVARSVIRRLINSQKSP
		56		0,596		MNQMVARSVIRRLINSQKSP
		57		0,597		MNQVVARSVIRRLINSQKSP

The SH method was used to detect putative functional residues amongst the identified AOX1 and 2 amino acid sequences. The method was applied as described in section 2.2.7. The results are summarized in Table 3.7 and represented graphically in the deduced AOX sequences in Appendix 8. In order to improve the understanding of the residues functionality indicated in the present study, the residues already referred in literature as having a functional importance were also represented in the same appendix

A total of 102 residues were indicated as potentially of functional relevance in the evaluated AOX sequences. Residues found from positions 2 to 65 will affect mainly the TP and not the protein functionally. In this regard, a group of residues occupying positions 24 to 61, requires special attention, because in some sequences these residues are part of the protein itself and not of the TP. The remaining residues sites must be carefully analyzed with regard to their position in the sequence and the residues with known functional importance around them. For the establishment of functional relevance, additional research will be required for each specific residue. As indication, the residues occupying positions located in regions already tested and proved functional, are the most promising ones. A graphical representation of the residues is provided in Appendix 8 to facilitate inspection.

Table 3.7 - Putative functional residue sites of the aligned AOX amino acid sequences identified using the SH method (see section 2.2.7). The positions indicated are represented in the alignment presented in Appendix 8. **A** - AOX1 subgroup; **B** - AOX2 subgroup; The '**Consensus**' columns give all residue present in subgroups A and B, respectively, in order of decreasing frequency and in lowercase when the frequency is less than half (see Appendix 7 for residue name); **Grey lines** - Residues with known functional relevance (Appendix 8); **Position** - Residue position in the alignment; **Rank** - Number of neighboring sites below the cutoff.

Position	Entropy				SH cutoff	Rank	Consensus	
	A	B	AB	rel.			A	B
2	0.00	0.00	0.78	1.24	0.00	9	M	N
3	0.00	0.96	1.52	1.24	0.00	9	M	HQ
4	0.00	1.33	1.80	1.24	0.00	9	T	Lmv
5	0.00	0.96	1.52	1.24	0.00	9	R	LV
6	0.00	0.00	0.78	1.24	0.00	9	G	A
7	0.00	0.96	1.52	1.24	0.00	9	T	KR
9	0.00	0.00	0.78	1.24	0.00	9	R	V
10	0.00	0.96	1.52	1.24	0.00	9	V	MI
11	0.00	0.00	0.78	1.24	0.00	9	A	R
14	0.00	0.00	0.78	1.24	0.00	2	T	I
15	0.99	0.96	1.76	1.24	0.00	2	MT	SN
24	0.00	0.18	0.93	1.24	0.00	9	-	Sr
25	0.00	0.96	1.52	1.24	0.00	9	-	PQ
26	0.00	0.96	1.52	1.24	0.00	9	-	AK
27	0.00	0.96	1.52	1.24	0.00	9	-	PS
28	0.00	1.13	1.65	1.24	0.00	9	-	SPa
29	0.00	1.26	1.75	1.24	0.00	9	-	Lim
30	0.00	0.96	1.52	1.24	0.00	9	R	TS
31	0.00	0.96	1.52	1.24	0.00	9	L	IT
33	0.00	0.00	0.78	1.24	0.00	9	S	R
36	0.00	0.96	1.52	1.24	0.00	21	K	TD
37	0.00	0.99	1.54	1.24	0.00	21	G	DE
38	0.00	0.96	1.52	1.24	0.00	21	-	SI
40	0.00	0.96	1.52	1.24	0.00	21	A	TI
42	0.00	0.96	1.52	1.24	0.00	21	E	RN
43	0.95	0.00	1.01	1.24	0.00	21	SG	R
45	0.00	0.96	1.52	1.24	0.00	21	K	SR
46	0.00	0.00	0.78	1.24	0.00	21	F	L
47	0.00	0.96	1.52	1.24	0.00	21	P	VG
48	0.00	0.96	1.52	1.24	0.00	21	V	YI
49	0.00	0.96	1.52	1.24	0.00	21	M	VI
51	0.00	0.00	0.78	1.24	0.00	21	-	G
52	0.00	0.00	0.78	1.24	0.00	21	-	G
53	0.00	0.96	1.52	1.24	0.00	21	-	GA
54	0.00	0.96	1.52	1.24	0.00	21	-	VR
55	0.00	0.96	1.52	1.24	0.00	21	-	EV
56	0.00	0.65	1.28	1.24	0.00	21	-	Lf
57	0.00	0.96	1.52	1.24	0.00	21	-	MG
58	0.00	0.89	1.47	1.24	0.00	21	V	Kt
60	0.00	0.00	0.78	1.24	0.00	21	W	M
61	0.00	0.00	0.78	1.24	0.00	21	R	M
64	0.00	0.96	1.52	1.24	0.00	3	L	E-
65	0.00	0.00	0.78	1.24	0.00	3	T	A
66	0.00	1.33	1.80	1.24	0.00	3	L	Vae
69	0.00	1.26	1.75	1.24	0.00	1	K	Tnd
72	0.00	1.26	1.75	1.24	0.00	2	V	Kas
73	0.00	0.00	0.78	1.24	0.00	2	N	K
84	0.44	0.96	1.63	1.24	0.00	7	Nd	K-
85	0.44	0.96	1.63	1.24	0.00	7	Kn	E-
86	0.44	0.96	1.63	1.24	0.00	7	Nk	E-
87	1.24	1.35	2.11	1.24	0.00	7	Gsr	-Ek
88	0.00	0.96	1.52	1.24	0.00	7	E	K-

Table 3.7 - *Continued*

Position	Entropy				SH cutoff	Rank	Consensus	
	A	B	AB	rel.			A	B
89	0.00	0.96	1.52	1.24	0.00	7	D	K-
91	0.95	0.00	1.01	1.24	0.00	7	G-	E
94	0.95	0.00	1.01	1.24	0.00	1	EQ	V
97	0.00	0.96	1.52	1.24	0.00	1	A	ST
102	0.00	0.00	0.78	1.24	0.00	5	I	V
103	0.00	0.00	0.78	1.24	0.00	5	K	A
104	0.00	0.00	0.78	1.24	0.00	5	G	R
105	0.44	0.00	0.89	1.24	0.00	5	Eq	P
106	0.00	0.96	1.52	1.24	0.00	5	E	RK
114	0.00	0.96	1.52	1.24	0.00	2	P	ED
116	0.00	0.00	0.78	1.24	0.00	2	K	P
121	0.00	0.00	0.78	1.24	0.00	1	R	M
131	0.00	0.00	0.78	1.24	0.00	1	T	S
141	0.00	0.00	0.78	1.24	0.00	1	T	K
147	0.00	0.89	1.47	1.24	0.00	1	L	Vm
150	0.00	0.65	1.28	1.24	0.00	1	W	Kr
154	0.00	0.89	1.47	1.24	0.00	1	S	Li
157	0.00	0.96	1.52	1.24	0.00	1	F	IL
191	0.00	0.00	0.78	1.24	0.00	1	C	L
196	0.00	0.00	0.78	1.24	0.00	2	R	K
198	0.00	0.00	0.78	1.24	0.00	2	E	Q
206	0.00	0.00	0.78	1.24	0.00	1	T	A
209	0.00	0.00	0.78	1.24	0.00	1	D	E
221	0.00	0.00	0.78	1.24	0.00	4	F	M
222	0.00	0.00	0.78	1.24	0.00	4	M	V
224	0.00	0.00	0.78	1.24	0.00	4	V	L
225	0.00	0.00	0.78	1.24	0.00	4	S	V
228	0.00	0.00	0.78	1.24	0.00	1	R	K
233	0.00	0.96	1.52	1.24	0.00	1	A	FL
236	0.00	0.00	0.78	1.24	0.00	1	F	L
246	0.00	0.00	0.78	1.24	0.00	5	Y	F
248	0.00	0.00	0.78	1.24	0.00	5	L	V
249	0.00	0.00	0.78	1.24	0.00	5	A	L
251	0.00	0.96	1.52	1.24	0.00	5	L	MI
252	0.00	1.13	1.65	1.24	0.00	5	A	MLv
275	0.00	0.00	0.78	1.24	0.00	1	F	Y
278	0.00	0.00	0.78	1.24	0.00	4	E	D
279	0.00	0.00	0.78	1.24	0.00	4	L	I
281	0.00	0.96	1.52	1.24	0.00	4	K	SR
283	0.00	0.96	1.52	1.24	0.00	4	T	AL
301	0.00	0.96	1.52	1.24	0.00	3	A	KQ
303	0.00	0.00	0.78	1.24	0.00	3	S	A
304	0.00	0.00	0.78	1.24	0.00	3	T	K
309	0.00	0.00	0.78	1.24	0.00	3	V	I
310	0.44	0.96	1.63	1.24	0.00	3	Mi	TL
312	0.00	0.00	0.78	1.24	0.00	3	V	I
331	0.00	0.00	0.78	1.24	0.00	1	Y	F
334	0.00	0.00	0.78	1.24	0.00	1	H	K
337	0.00	0.00	0.78	1.24	0.00	2	K	R
339	0.00	0.00	0.78	1.24	0.00	2	S	A

The residues positions indicated in grey were found at regions or positions already proven to affect the protein functionality. Nevertheless, in order to understand its importance in the sequences, additional work must be performed.

3.4.2 - Analysis of *AOX* non-coding region

Aiming to understand the variability found at the *AOX* intron level and searching for a possible correlation with the detected SE efficiency phenotype for each cell line, clustering analyses were performed, comparing all intron 1, 2 and 3 sequences, each on separated analyses (Figures 3.13, 3.14 and 3.15).

The intron 1 sequences from each of the evaluated genes cluster together with the sequences from the same gene. Interestingly, the *AOX2a* intron 1 sequences, are more related with the ones from the *AOX1*, than with the ones from *AOX2b*. The *AOX2a* intron 1 sequences from subspecies *gadecaei* and *gummifer*, form a

separated cluster, as well as the one from the subspecies *halophilus*. The sequences L142 and L128 from the cultivar 'Rotin', also cluster together with the *halophilus* sequences. The remaining cultivars *AOX2a* intron 1 sequences, cluster together in a third group.

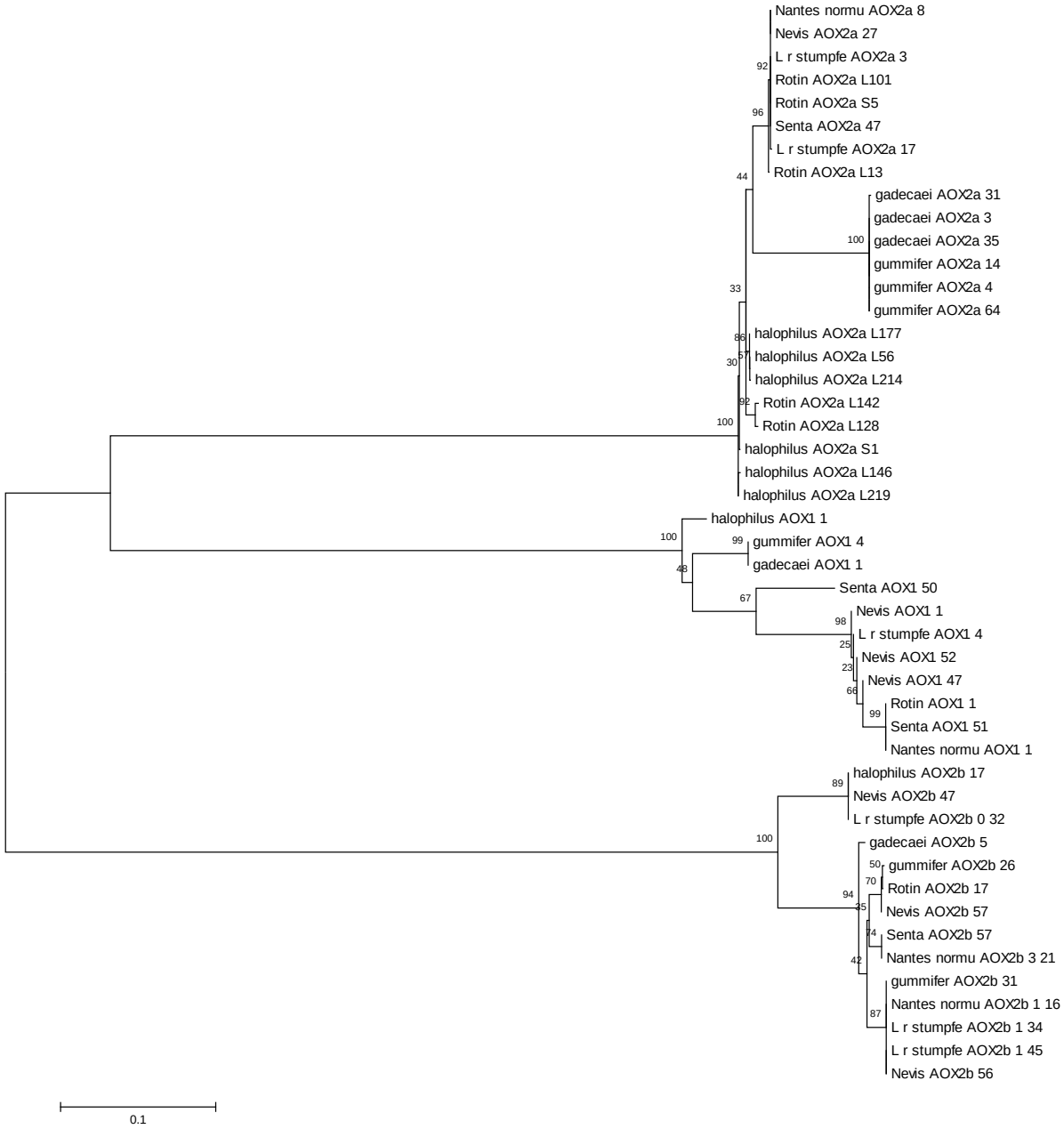


Figure 3.13 - Clustering analyses of the intron 1 from the 47 *AOX1*, *2a* and *2b* nucleotide sequences identified in the 8 cell lines. The clustering analyses were performed using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the analyses. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 2476 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

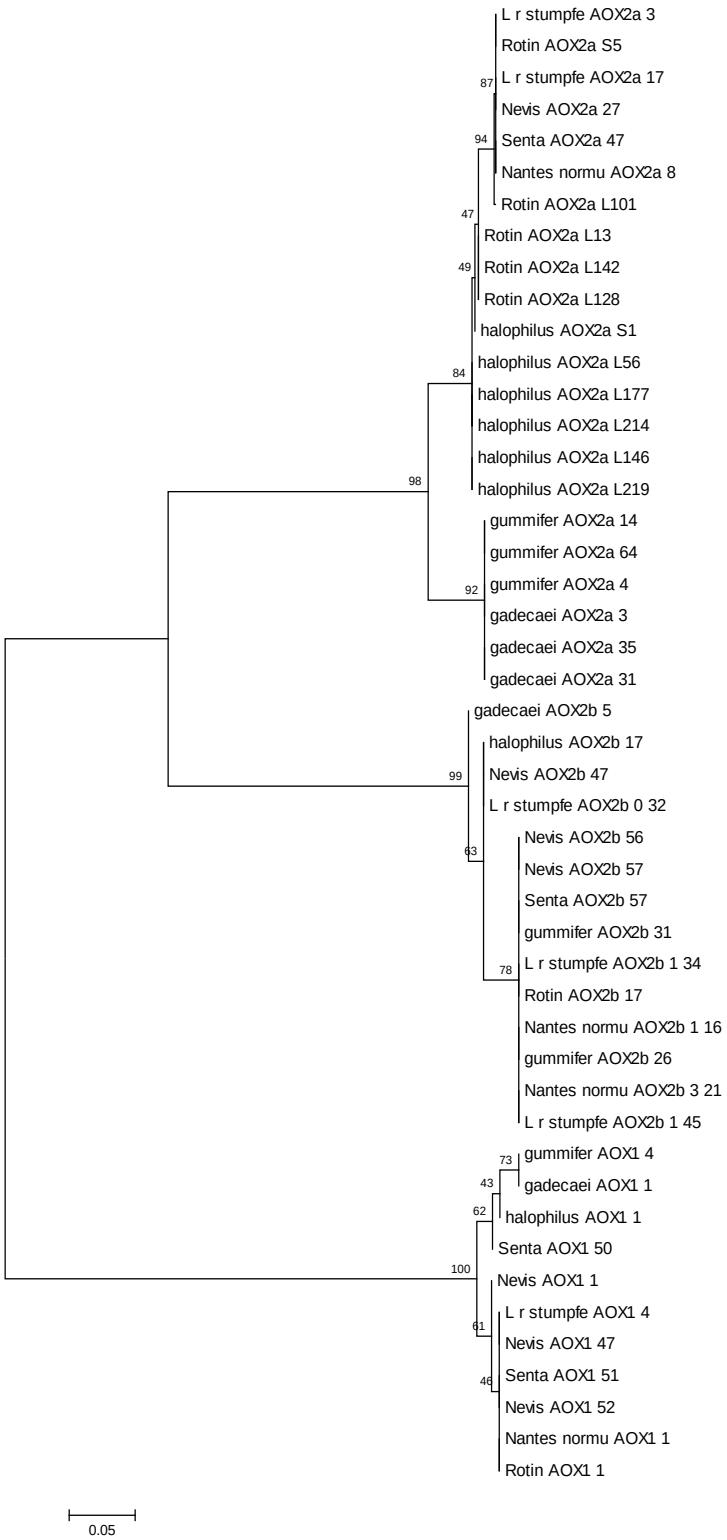


Figure 3.14 - Clustering analyses of the intron 2 from the 47 *AOX1*, *2a* and *2b* nucleotide sequences identified in the 8 cell lines. The clustering analyses were performed using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the analyses (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 960 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

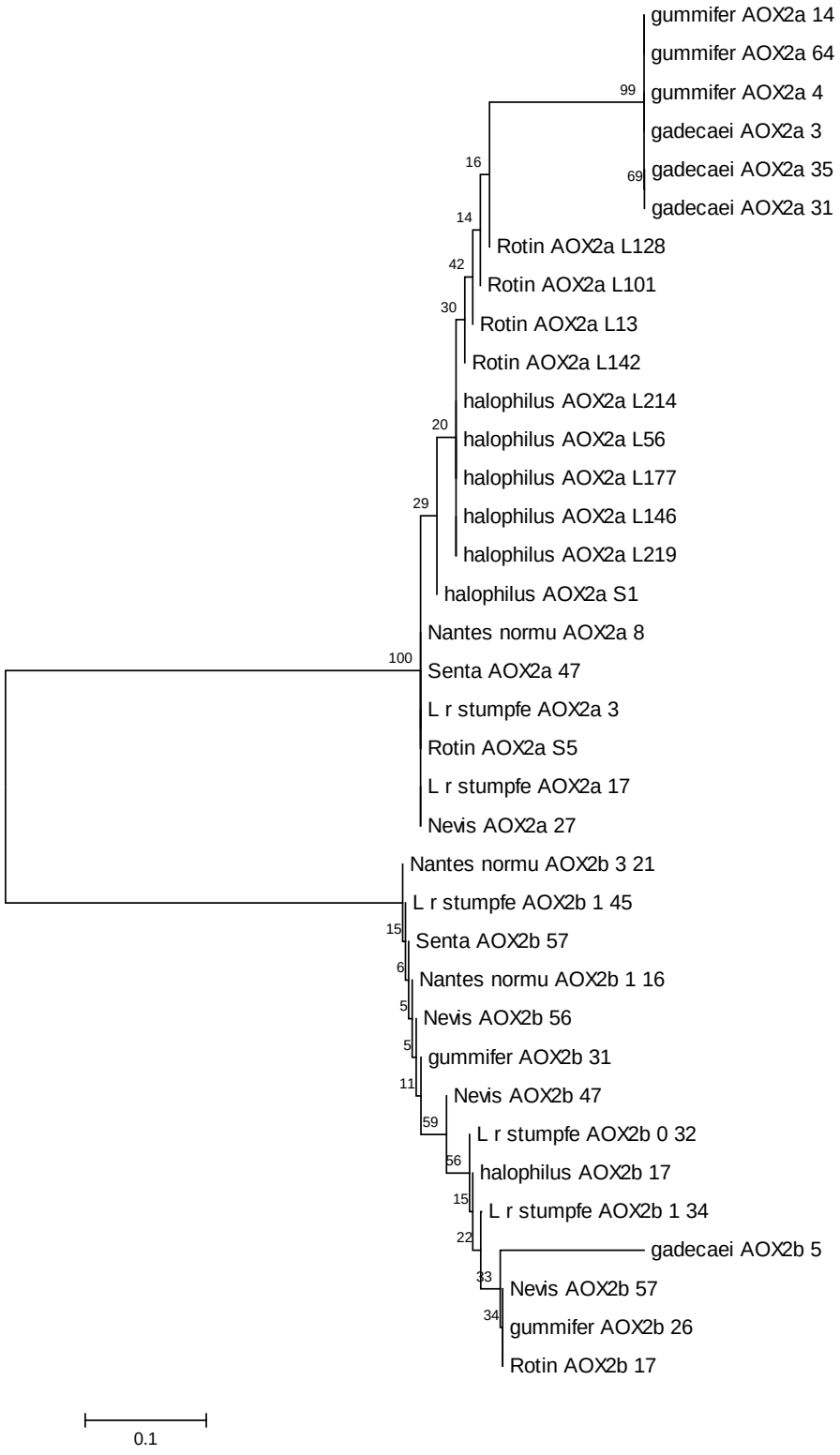


Figure 3.15 - Clustering analyses of the intron 3 from the 39 *AOX2a* and *2b* nucleotide sequences identified in the 8 cell lines. The clustering analyses were performed using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the analyses (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the analyses used to infer the phylogenetic tree. The clustering analyses distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 1318 positions in the final dataset. Clustering analyses were conducted in MEGA5 (Tamura *et al.*, 2011). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'.

AOX1 intron 1 sequences also cluster all together in a separated group. However, a clear separation among the sequences from subspecies and the ones from cultivars, can be observed, with the exception of sequence 50 from cultivar 'Senta', which appears isolated. In the case of *AOX2b* intron 1 sequences, two clusters were identified. The first grouping sequences from the subspecies *halophilus* and from the cultivars 'Nevis' and 'L r stumpfe'. The second cluster contains all the remaining sequences, which includes subspecies and cultivars, with no clear separation amongst them.

Just as with intron 1, also intron 2 sequences cluster together with the sequences from the same gene. *AOX2a* intron 2 sequences form two main clusters, one grouping the sequences from *gummifer* and *gadecaei* subspecies, the other grouping all the remaining ones. On this second cluster the *halophilus* sequences appear grouped together with the sequences L13, L142 and L128 from the cultivar 'Rotin'. The other sequences from the cultivars grouped all on a single group. The *AOX2b* intron 2 sequences form three different clusters. The first just contains the sequence 5 from the subspecies *gadecaei*, which seems to be a unique sequence. The second grouped the sequence from the *halophilus* subspecies (17) and a single sequence from the cultivar 'Nevis' (47) and 'L r stumpfe' (0_32). The third group contains all the remaining sequences from the cultivars and the subspecies.

Due to the fact that *AOX1* does not have intron 3, clustering analyses for intron 3 only includes *AOX2a* and *AOX2b* sequences. Likewise for intron 1 and intron 2 sequences, also intron 3 sequences clustered with the sequences belonging to the gene of origin. The *AOX2a* intron 3 sequences from the subspecies *gummifer* and *gadecaei* form a separated group. Also, the 'L' sequences from the cultivar 'Rotin' and the subspecies *halophilus* grouped separately, indicating the presence of a characterizing mutated region. The 'S' sequences, also from 'Rotin' and *halophilus* accessions form the remaining group containing the sequences from all other cultivars.

On the other hand, the *AOX2b* intron 3 sequences cluster all mixed, subspecies and cultivars, with no clear distinction amongst accessions.

A search for enrichment of functional *cis*-acting regulatory elements was performed in the highly polymorphic intron 1 sequences, similarly to the search performed on *AOX* ORFs sequences. The search was focused on intron 1, because it is frequently the promoter proximal intron the one with improved capacity to concentrate signals affecting gene expression and transcription (Rose *et al.*, 2008; Parra *et al.*, 2011; Gallegos and Rose, 2015). In the same way as performed for ORFs sequences, selected polymorphic intron 1 was analyzed using two software (Nsite-PL and ScanWM-P). The resulting analysis is summarized in Table 3.8 and represented on Appendix 9 (9D - *AOX1*, 9E - *AOX2a* and 9F - *AOX2b*).

Table 3.8 - Enrichment of putative regulatory elements in the identified *AOX1*, *2a* and *2b* highly polymorphic intron 1 from the 8 selected cell lines assessed using Nsite-PL and ScanWM-P software (see section 2.2.7). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed (see Appendix 3); **Grey cells** - Sequences with different levels of identified regulatory elements/weight matrices/motifs enrichment within accessions; **Nr.** - Number.

				Nsite-PL (Nr.)		ScanWM-P (Nr.)	
	Cell line	Clone	Length (bp)	Regulatory elements	Motifs	Weight Matrixes	Motifs
AOX1	**2_4**	**50**	***630***	23	33	12	14
		51	***712***	29	29	5	5
	4_5	**1**	***712***	29	29	12	14
	8_2	**4**	***707***	34	34	12	10
	18_1	**1**	***636***	38	57	5	10
	19_1	**4**	***707***	34	44	12	8
	20_2	**1**	***707***	34	44	12	8
	26_L5.S.R.	**1**	***712***	29	29	12	14
	27_2	**1**	***707***	34	34	12	10
		47	***700***	31	31	12	10
		52	***707***	34	34	12	10
AOX2a	**2_4**	**47**	***2071***	25	28	32	52
	4_5	**8**	***2070***	25	28	32	52
	8_2	**3**	***2070***	25	28	32	52
		17	***2068***	25	27	32	52
	18_1	**S1**	***2076***	28	32	31	50
		L56	***1981***	25	40	33	53
		L146	***1957***	29	33	32	50
		L177	***1979***	25	38	33	53
		L214	***1980***	25	38	33	53
		L219	***1957***	27	31	32	50
	19_1	**4**	***1987***	23	32	34	70
		14	***1989***	23	34	34	70
		64	***1985***	23	30	34	70
	20_2	**3**	***1987***	23	34	34	70
		31	***1984***	23	30	34	70
		35	***1985***	23	32	34	70
	26_L5.S.R.	**S5**	***2070***	25	28	32	52
		L13	***2070***	26	29	32	52
		L101	***2070***	25	28	32	52
		L128	***2061***	24	27	31	53
		L142	***2079***	24	26	33	55
	27_2	**27**	***2070***	25	28	32	52
AOX2b	**2_4**	**57**	***1019***	22	27	16	32
	4_5	**1_16**	***823***	26	36	16	31
		3_21	***1019***	22	27	16	32
	8_2	**0_32**	***560***	31	32	20	23
		1_34	***823***	26	36	16	31
		1_45	***823***	26	36	16	31
	18_1	**17**	***560***	32	31	20	23
	19_1	**2_26**	***823***	22	31	13	27
		2_31	***823***	26	36	16	31
	20_2	**5**	***832***	23	33	17	31
	26_L5.S.R.	**17**	***822***	22	31	13	27
	27_2	**47**	***560***	31	32	20	23
		56	***823***	26	36	16	31
		57	***823***	22	31	13	27

The Nsite-PL software detected differences in the *AOX1* intron 1 sequences within accessions (2_4 and 27_2) as well as amongst the remaining accessions sequences. In the same way, also ScanWM-P detected differences, however, with a low level of enrichment at the level of regulatory elements as well as at the level of detected motifs, being unable to detect differences amongst the 27_2 accession variant sequences. The analysis of *AOX1* intron 1 sequences (Appendix 9D), reveals that Nsite-PL detected regulatory elements on 61 polymorphic positions and ScanWM-P on 25. When considering the results obtained with the Nsite-PL, the sequence 50 from the cultivar 'Senta' alone has 13 specific positions occupied by regulatory elements, being the most specific *AOX1* sequence, followed by the *halophilus* sequence with seven specific positions. The W is the regulatory element found enriched when the Nsite-PL was used. When ScanWM-P results are considered, also the 'Senta' sequence 50 and the one from the subspecies *halophilus* were the most specific, with three unique positions each occupied by regulatory elements. In the same way as for the Nsite-PL, also ScanWM-P found the W regulatory element enriched in the *AOX1* sequences evaluated. The W usually is the binding domain for WRKY transcription factors. However, the detected regulatory elements in the *AOX1* were dispersed along the sequences, with no clear distinction amongst the identified SE efficiency phenotypes.

In the case of *AOX2a* intron 1, the Nsite-PL and ScanWM-P software identified the highest level of enrichment on the variant sequences identified for accessions 18_1 and 26_L5.S.R.. From both software used, only Nsite-PL was able to identify differences amongst variant sequences of *AOX2a* intron 1 accessions 8_2, 19_1 and 20_2. With the analysis of *AOX2a* intron 1 sequences (Appendix 9E), the Nsite-PL detected 34 polymorphic positions occupied by regulatory elements, while 79 were detected when the ScanWM-P was used. When the Nsite-PL results are considered, just the sequence L146 from the subspecies *halophilus* had a specific position occupied by a regulatory element, indicated by the number 19 in the Appendix 9E. In the case of ScanWM-P, three sequences were reported with a single specific position each one, namely: the L142 from 'Rotin' (regulatory element number 40), the L56 from *halophilus* (regulatory element number 60) and finally the 14 from *gummifer* (regulatory element number 71). When the Nsite-PL *AOX2A* results were considered the regulatory elements GA motif (GA5) and the CT-rich were found repeatedly in the

sequences, with several levels of enrichment depending on the sequence (see Appendix 9E). These regulatory elements (indicated by the numbers three, four and six in the Appendix 9E) were found several times in a row, as happened in the case of the sequence 3 from the accession *gadecaei*. The biological meaning of this intron enrichment remains unknown. When the ScanWM-P results were analyzed, the number of elements found repeatedly in the *AOX2a* sequences was higher. It was possible to find the regulatory elements A1, W, ATCATC motif, ABRE's and G box (G) along the sequences several times at different positions. This analysis performed with the ScanWM-P was the one that produced the largest number of hits, with a total of 79 polymorphic positions occupied by regulatory elements. However, the level of enrichment was also higher.

For *AOX2b* intron 1, differences were detected amongst accessions and variant sequences within accessions in the same manner, using both software. Differences on motif enrichment were able to differentiate variant sequences on accessions 4_5, 8_2, 19_1 and 27_2. With the analysis of the *AOX2b* intron 1 sequences (Appendix 9F), the Nsite-PL detected 40 polymorphic positions occupied by regulatory elements, while the ScanWM-P detected 33. When the Nsite-PL results were considered, none of the sequences had a specific position occupied by regulatory elements, but in the case of the ScanWM-P results, the sequence from accession *gadecaei* was found with three specific polymorphic positions occupied by regulatory elements (indicated in the Appendix 9F with the numbers 7, 9 and 28).
When considering the Nsite-PL results, the barley H21 element from the SyntheticOLIGOs gene (H21), AC-I, TFIIIA-type zinc finger motif from petunia (ZPT2-2) and Wuschel 2 (WUS2) were found enriched in the sequences. On the other hand, when ScanWM-P results were considered, the elements found enriched in the sequences were the opaque-2d binding site (O2d), W, ATCATC, TGA1 and the auxin response (AuxRE). The regulatory elements identified at polymorphic positions were distributed along the sequences, also with no clear distinction amongst the SE efficiency phenotype.

In order to understand the putative intron mediated-enhancement of gene expression of the *AOX*s intron 1 sequences identified in the eight cell lines with extreme SE efficiency phenotypes, and the possible correlation with the detected phenotype, the

IMEter V2.1 software was used, according to the defined strategy described in the section 2.2.7 (Table 3.9).

The results obtained from the IMEter software reveal that *AOX1* intron 1 sequences are the ones with the lowest capacity to increase the gene expression, and the *AOX2a* sequences the ones with the higher. The IMEter scores obtained for the different *AOX1* and *AOX2a* intron 1 sequences do not present any correlation with the detected SE efficiency phenotype, neither, when the forward or the reverse strand were evaluated for the signal presence.

Table 3.9 - Assessment of mediated-enhancement by *AOX1*, *2a* and *2b* intron 1 sequences using IMEter V2.1 (see section 2.2.7). **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed; **Grey cells** - Sequences from non-embryogenic cell lines.

			Forward strand		Reverse strand	
	Cell line	Clone	IMEter score	Percentile	IMEter score	Percentile
AOX1	**2_4**	**50**	8.41	91	6.27	86
		51	8.85	92	8.18	91
	4_5	**1**	8.85	92	8.18	81
	8_2	**4**	7.85	90	5.46	82
	18_1	**1**	9.32	93	6.64	87
	19_1	**4**	6.71	87	10.08	94
	20_2	**1**	6.71	87	10.08	94
	26_L5.S.R.	**1**	8.85	92	8.18	91
	27_2	**1**	7.85	90	5.46	82
		47	6.74	87	4.29	75
		52	7.85	90	5.46	82
AOX2a	**2_4**	**47**	12.96	96	14.63	97
	4_5	**8**	12.96	96	14.63	97
	8_2	**3**	12.96	96	14.63	97
		17	12.98	96	14.32	97
	18_1	**S1**	11.45	95	13.71	97
		L56	10.88	94	16.11	98
		L146	10.92	94	15.71	97
		L177	10.90	94	15.80	97
		L214	11.45	95	14.12	97
		L219	11.45	95	13.71	97
	19_1	**4**	11.06	95	13.46	96
		14	11.09	95	13.06	96
		64	11.12	95	12.66	96
	20_2	**3**	11.06	95	13.46	96
		31	11.09	95	13.06	96
		35	11.12	95	12.66	96
	26_L5.S.R.	**S5**	12.96	96	14.63	97
		L13	12.96	96	14.62	97
		L101	12.96	96	14.63	97
		L128	11.50	95	13.32	96
		L142	11.86	95	14.18	97
	27_2	**27**	12.96	96	14.63	97
AOX2b	**2_4**	**57**	5.64	83	8.51	92
	4_5	**1_16**	8.25	91	11.59	95
		3_21	5.64	83	8.51	92
	8_2	**0_32**	5.49	82	8.77	92
		1_34	8.25	91	11.59	95
		1_45	8.25	91	11.59	95
	18_1	**17**	5.49	82	8.77	92
	19_1	**2_26**	8.25	91	11.59	95
		2_31	11.26	95	14.59	97
	20_2	**5**	11.43	95	14.05	97
	26_L5.S.R.	**17**	11.28	95	14.60	97
	27_2	**47**	5.49	82	8.77	92
		56	8.25	91	11.59	95
		57	11.26	95	14.59	97

Interestingly, when the *AOX2b* intron 1 sequences were evaluated, the results showed differences which could help on that correlation. Almost all sequences belonging to embryogenic cell lines present higher IMEter scores in the forward and in the reverse strand. The exceptions to this observation are the scores obtained by the sequences from accessions 'Nantes normu' and '*gummifer*', both non-embryogenic, but also with high IMEter scores.

Although not clearly correlated, the indications obtained with the IMEter results for the *AOX2b* intron 1 sequences, seems to suggest that the gene may be associated with the capacity to develop embryos in *Daucus*. However, the results should be regarded carefully and require further investigation in order to support this view.

DISCUSSION

4 - Discussion

The identification by Costa *et al.* (2009 and 2014b) of three *AOX* genes in the genome of *D. carota* and the works from Frederico *et al.* (2009a) and Zavattieri *et al.* (2010), allowed to establish a possible link between the polymorphism occurrence in the *AOX* genes and the SE response efficiency of *Daucus* explants.

Aiming on continuing the efforts to understand and clarify this possible relation, the first priority of the present study was to establish a *Daucus* cell line collection with different SE efficiency phenotypes, using an innovative phenotyping approach for cell line evaluation and selection. To the best of our knowledge, this represents the largest study performed so far, in order to compare *Daucus* SE efficiency, among and within accessions. *Daucus* has been widely used for the study of the network of biological processes occurring during SE and allowed the discovery of important elements associated with the onset of SE in plants (Satoh *et al.*, 1986; Thomas *et al.*, 1989; Schmidt *et al.*, 1997; Kikuchi *et al.*, 2006 and references therein), being now considered as a model species for SE studies.

Previous studies looking for cell line efficiency phenotyping, frequently considered a single accession (Takahata, 2008; Shibukawa *et al.*, 2009), or a single cell line (Pennell *et al.*, 1992) and, limited attention was given to differences among accessions, or cell lines derived from a single accession, in order to obtain broad conclusions.

The innovative two stages phenotyping approach developed here, allowed to select cell lines with stable SE efficiency, over replicates and time, supporting it as a good procedure for phenotyping SE cell line efficiency and stability, whenever a large number of accessions are considered.

Furthermore, common methods used for phenotyping SE cell line efficiency are primarily based on embryo counting (Tangolar *et al.*, 2008; Naing *et al.*, 2013). This task is impossible to employ when a large number of accessions or cell lines are used, as is the case in the present study, which started with 28 accessions and 139 cell lines. This has been probably the reason why most of SE studies published so

far only consider a few or a single cell line/accession. This restriction of the methodology mostly applied for this issue and the ways to overcome it, have been largely discussed by Fiorani and Schurr (2013) and duly taken in consideration in the present study. The approach here developed, considers growth dynamics over time and imaging technologies, to overcome the phenotyping bottleneck limitation imposed by embryo counting, which is usually found when a large number of accessions must be screened for SE efficiency selection.

By applying the proposed methodology, from the 28 accessions initially evaluated, three were unable to develop *calli*, and were not considered, namely: *D. c pusillus*, *D. c. littorallis* and *D. c. muricatus*. The suspecies *pusillus* and *muricatus* were previously classified as non-embryogenic by Imani *et al.* (2001) and Thi and Pleschka (2005), which seems to confirm our results. The same authors also refer *D. montevidensis* as non-embryogenic. In our study, it presents embryogenic capacity, although showing an unstable phenotype behavior. Recently, also Tavares *et al.* (2010) classified *D. c. halophilus* as embryogenic, but without considering differences within the accession. Our results confirmed this capacity, but two from the six *D. c. halophilus* cell lines evaluated here were classified as non-embryogenic. In conclusion, at the end of the first stage of SE phenotyping selection, from the 25 *Daucus* accessions developing *calli* and, from which it was possible to establish true-to-type cell lines, all were reported as embryogenic at least in one of the cell lines tested. However, concerning the efficiency and stability of the SE phenotype, differences among and within accessions derived cell lines were detected when replicates were performed, partly confirming the previous results obtained by Imani *et al*. (2001) and Thi and Pleschka (2005), which claimed these differences, but only when different accessions had been used.

In order to improve the cell lines characterization at the end of the first phenotyping set, in terms of genetic diversity and phylogenetic relationships, the selected cell lines were further evaluated using the cTBP method. This method describes a successful and widely applicable ILP-based marker approach, that takes into account the amplification of either of the two introns, commonly present in conservative positions within the coding sequences of plant β-tubulin genes (Breviario *et al.*, 2007; Braglia *et al.*, 2010). In this view, the method has been successfully used for

fingerprinting the genomes of several plant species including those for which no genomic information was available (Braglia *et al.*, 2010). The cTBP evaluation allowed to assess cell lines genetic diversity and to identify those with an identical tubulin genetic background, giving as well an overview concerning the phylogenetic relationships amongst them. Interestingly, from the ones with identical cTBP profile, some presented a divergent SE efficiency phenotype and were selected for a second SE phenotype evaluation. Likewise, some indications concerning the breeding history of some cultivars, as in the case of 'Senta', was possible to infer from the analysis, in the sight of the *Daucus* domestication and breeding material origin clarification as stated and reviewed by Grzebelus *et al.* (2011 and 2014).

For the second step of SE efficiency evaluation, cell lines were selected based on the first phenotyping set results as well as on cTBP evaluation. By selecting 22 cell lines, belonging to 17 *Daucus* accessions of the 25 initially used, it was intended to cover as maximum as possible the genetic diversity presented in the study. The *NE* phenotype proved to be the most stable across the study, followed by the *VE* phenotype. At the first phenotyping set end, 11 cell lines expressed the *NE* phenotype in a stable way, against the seven displaying the *VE*. During the re-evaluation performed in the second set, as *NE*, six cell lines were selected, and five maintained the phenotype at the end. On the other hand, as *VE*, five cell lines were selected, but only two maintained identical phenotype classification. In general, the embryogenic phenotypes were the most unstable and the reason for this behavior may be related to the complexity of networks of the biological processes evolved during SE (Zeng *et al.*, 2007; Yang and Zhang, 2010; Smertenko and Bozhkov, 2014; Mahdavi-Darvari *et al.*, 2015; De-la-Peña *et al.*, 2015). SE has always been reported as an unstable process, highly variable over time, limiting its application in modern plant breeding (Deo *et al.*, 2010), although highly desirable in the view of mass multiplication up-scaling (Sujatha, 2011). In any way, and taking the previously stated, at the end of the second phenotyping set it was possible to identify several cell lines with stabilized SE efficiency phenotype across both sets of replicates and over time. Of those, eight were phenotyped for the most extreme SE efficiency phenotypes established (*VE*, *VE/E* and *NE*), being selected as the basic collection for genomic studies concerning *AOX* polymorphism evaluation.

A limitation usually associated with the selection and establishment of cell lines is the development of polyploidy (Ronchi *et al.*, 1992; Kubalakova *et al.*, 1996; Endemann *et al.*, 2001; Ishigaki *et al.*, 2014). Polyploid development has also been associated with the cell lines incapacity to develop somatic embryos (Coutos-Thevenot *et al.*, 1990; Konieczny *et al.*, 2012). On our trials, of the three cell lines developing polyploidy, two were found to be highly embryogenic and one was non-embryogenic, behavior that does not match with known bibliographic statements. It should however be taken in consideration that, the present study was focused on the capacity of the cell lines to undergo cell reprogramming events, leading to efficient SE, and not on the ability of embryos to develop into fully functional plants. This was the reason why ploidy was not evaluated at the embryo level. Besides, as addressed by Ronchi *et al.* (1992), the possibility that regenerated embryos from the polyploid lines were diploid remains possible, over the existence of polyploids, through mechanisms alternative to mitosis (reductional grouping or prophase chromosome reduction). Nevertheless, the remaining cell lines with normal ploidy levels were found to be embryogenic and non-embryogenic, which raised the question, if really polyploidy was the cause of SE recalcitrance as reported in literature. In line with the results acquired here with *Daucus*, also Zhang *et al*. (2006) was unable to establish a correlation between the incompetence for the SE and the development of polyploidy using *Citrus calli*. Held together all the above doubts and the ones named by Bennett (2004), Sun *et al.* (2011) and Yildiz (2013), where polyploidy is often linked to improved fitness characteristics and increased adaptability and tolerance to adverse environmental conditions, it was determined to continue the study using also the polyploid cell lines, which could give additional information concerning the *AOX* genomic regions prone to mutation events.

The final cell lines collection used for *AOX* molecular evaluation was obtained based on the results from SE efficiency, cTBP and ploidy characterization. This work represents the most comprehensive *AOX* variability assessment performed so far and may represent a step onward in the elucidation of AOX involvement on plant abiotic stress reactions, represented by the SE process in the present study, as discussed by Frederico *et al.* (2009a), Zavattieri *et al.* (2010) and Grafi *et al.* (2011). Using eight cell lines (four embryogenic and four non-embryogenic) derived from an identical number of *Daucus* accessions, 47 complete sequences (from start to end

codon), comprising all three genes found on *Daucus* (Costa *et al.,* 2009 and 2014b), were identified as unique from the 290 initially amplified using PCR technology and its variability evaluated using bioinformatic tools. From those, 11 belong to *AOX1*, 22 to *AOX2a* and 14 to *AOX2b*. Of the three genes amplified, *AOX1* was the one with the lowest level of sequence variability found, with just two cell lines carrying more than a unique sequence. On the other hand, *AOX2a* was the one with the highest degree of sequence variation found, with six cell lines carrying more than a single sequence. *AOX2b* has four cell lines with more than a single sequence. Sequence variability was found associated indifferently with polyploid and diploid cell lines, which remains an intriguing event. Likewise, we cannot leave out the probability that not all variability was covered in the study, because the study concentrated on the gene regions in between the start and end codons, and just exons and introns were evaluated

Sequence variability at different *AOX* regions had already been reported in several plant species, including *Hypericum perforatum* (Ferreira *et al.*, 2009), *Olea europaea* (Santos-Macedo *et al.,* 2009), *Pinus pinea* (Frederico *et al.*, 2009b), or *Daucus.* (Cardoso *et al.*, 2009; Cardoso *et al.*, 2011). This suggests that *AOX* may be under strong environmental pressure and may be a solid candidate for functional marker development for abiotic stress, as previously reported by Arnholdt-Schmitt *et al.* (2006) and fulfilling the initial prerequisites indicated by Andersen and Lübberstedt (2003). In the present assessment, SNPs and InDels were the source of sequence variability and were strongly found at the level of intron sequences. However, variability was also found at the exon level, on *AOX1* and *AOX2a*. Three amino acid size variant sequences were found for AOX1 and two for AOX2a. Interestingly, the InDels found in exon 1 of *AOX1* lead to the existence of two AOX1 amino acid sequences within the *Daucus* accession 'Senta', and a 3 residues shorter one for the accessions 'L_r_stumpfe' and 'Nevis'. On *AOX2a*, size variability was detected, also in exon 1, in the accessions '*gummifer*' and '*gadecaei*', leading to a 1 residue shorter AOX2a amino acid sequence identification. AOX2b did not present any variability at the amino acid sequences size level. The variability found at the intron level, was extensive, in accordance with previous results from Cardoso *et al.* (2009) also working with *Daucus*. Those authors reported for individual plants an ILP in the intron 3 of *AOX2a* containing a repetitive deletion affecting a putative pre-micro RNA site,

allowing the grouping of genotypes. This ILP was also detected in the present study, associated with an embryogenic cell line (26_L5.S.R), as well as with a non-embryogenic one (18_1), which reduced its importance in the view of our goal. These cell lines were the ones with the highest level of *AOX2a* unique sequences detected, six in the case of the cell line 18_1 and five in the 26_L5.S.R. ILPs were also detected in the *AOX2b* intron 1 sequences, also, in accordance with the previously detected by Cardoso *et al*. (2011). However, ILPs at *AOX2b* intron 1 were detected in three of the eight cell lines evaluated, namely: 4_5, 8_2 and 27_2. The first was found non-embryogenic and others were embryogenic, with no clear relation to the ILP occurrence. ILP occurrence was not detected in *AOX1* sequences. On the other hand, data on gene expression obtained by Frederico *et al.* (2009a) and Campos *et al.* (2015), showed that *AOX1* and *AOX2a* are highly responsive during SE expression in *Daucus* using the cell line 26_L5.S.R., confirming its involvement during the process. Unfortunately, no data concerning *AOX2b* expression are available for *Daucus* during the SE developmental process, limiting a global overview concerning the gene activity during the process.

Multiple sequence alignments are often employed to reveal functionally important residues within a protein family and the development of algorithms able to identify key residues that determine functional differences between protein subfamilies, could be particularly useful (Capra and Singh, 2007). Likewise, conservation analysis has turned out to be a potent indicator of operational importance and has been applied to detect residues involved in ligand binding (Liang *et al.*, 2006), in protein-protein interaction interfaces (Caffrey *et al.*, 2004; Guharoy and Chakrabarti, 2005; Mintseris and Weng, 2005), in maintaining protein structure (Schueler-Furman and Baker, 2003), and in evaluation of protein functional specificity (Kalinina *et al.*, 2003). Conservation analysis has also been used in conjunction with structural information in many of these applications (Panchenko *et al.*, 2004; Landau *et al.*, 2005). Pirovano *et al.* (2006) and Feenstra *et al.* (2007) developed the so-called SH algorithm, using both AOX subfamilies protein sequences as an example for algorithm training. The SH method was employed using the AOX polymorphic amino acid sequences deduced at the present study, and 102 residues were shown as potentially important for regulation. At the protein level, and considering data previously reported, the AOX amino acid sequences deduced from our study, presented the diiron binding sites

conserved in accordance with the model proposed initially by Andersson and Nordlund (1999) and lately improved by Berthold *et al.* (2000). Also, the residue position identified by Frederico *et al.* (2009b) was found conserved in accordance with the model proposed. Some of the indicated residues by the SH analysis were found on regions already indicated by Crichton *et al.* (2005) to influence AOX regulatory behavior, as well as with the reference transmembrane helical regions (Berthold *et al.*, 2000; Saisho *et al.*, 2001; Heazlewood *et al.*, 2004). These findings may be helpful in understanding AOX differential activity amongst the gene subfamilies, as well as within gene variants. Interestingly, the biggest impact of the mutations (SNPs and InDels) with a visible effect on the amino acid sequence in terms of residue change, occurs at exon 1 level. According to the SH results, this mainly affects the TP, producing changes in its length without affecting the mitochondrial membrane processing. Cardoso *et al.* (2015), also stated that *AOX* exon 1 in plants was the main source of nsSNPs leading to residue changes in plants and our results confirm what was stated by those authors. Interestingly, this AOX region had been the least studied concerning the protein functionally, according to the literature assessment presented in Appendix 8. This observation reinforces the need for additional studies in order to clarify the protein activity in this neighborhood and in specific regions when residues differences amongst sequences had been detected.

Taking in consideration the few evidences concerning a correlation amongst the occurrence of *AOX* polymorphisms and the phenotyped SE efficiency of the selected cell lines, it was decided to initiate a search for possible regulatory elements present in the sequence. These elements may also be affected by invisible mutations, such as sSNPs, which do not produce any visible change at the amino acid sequence. This search was also motivated by recent reports (Rose *et al.*, 2008; Rose *et al.*, 2011; Parra *et al.*, 2011; Gallegos and Rose, 2015), where introns, in particular intron 1 was referred as a source of regulatory elements that may influence gene activities.

Aiming to identify such missing information in our data, a search for regulatory elements was performed at the *AOX* ORF and intron 1 nucleotide sequences, producing an extensive number of hits with known regulatory elements. Intriguingly a

GA motif (GA5), usually recognized by the basic pentacysteine 1 (BPC1) transcription factor, was found to be highly enriched in the *AOX2a* intron 1 sequences. BPC1 was identified as a regulator of the ovule identity gene *SEEDSTICK* in *Arabidopsis*, which is specifically expressed in ovules (Rounsley *et al.*, 1995; Pinyopich *et al.*, 2003; Brambilla *et al.*, 2007). BPC1 binds to the *SEEDSTICK* promoter at multiple GA-rich boxes (Kooiker *et al.*, 2005).

On the other hand, the results on *AOX2b* ORF regulatory elements give some strength to the possibility that the gene can be associated with embryo development efficiency. This is the case at least in some of the *Daucus* evaluated, with the finding of three regulatory elements (MSA, ABRE3 and Emb1) predominately present on embryogenic cell lines due to the existence of sSNPs. The MSA element had been usually referred in literature as associated with several relevant genes for the regulation of the cell cycle G2/M transition, including the cyclin-dependent serine/threonine kinases (Zhiponova *et al.*, 2006) and the MYB3R (Haga *et al.*, 2011). The ABRE3 belongs to the well known group of ABA responsive elements, usually found in ABA inducible genes, such as the ones encoding seed storage proteins, late embryogenesis abundant (LEA) proteins, and various other protein families (Zhang *et al.*, 2005). Importantly, ABA mediates many aspects of physiological responses to environmental stress, such as drought, cold and salinity. Many experiments have shown that abiotic stress also activates processes underlying requiring ABA signaling (Finkelstein *et al.*, 2002; Shinozaki and Yamaguchi-Shinozaki, 2000; Zhu, 2002). Specifically, a large number of genes that respond to abiotic stress are also inducible directly by ABA treatment (Seki *et al.*, 2002a and b), providing direct evidence that ABA must be involved in the processes responding to these environmental stresses. The early methionine (Em) protein from a number of plants accumulates to high levels exclusively in the embryo during the maturation stage of seed development (Schultz *et al.*, 1996). The Emb1 element was firstly described associated with the gene encoding the Em protein in wheat (Marcotte *et al.*, 1989) and later identified as part of the ABA-response network of several genes during stress induced responses (Uno *et al.*, 2000; Narusaka *et al.*, 2003; Shen *et al*., 2004). Still, the elements here reported had been associated with the promoters and not with coding regions as in the *AOX2b* case. Therefore, the biological meaning of those elements occurrence in the *AOX2b* exon associated with

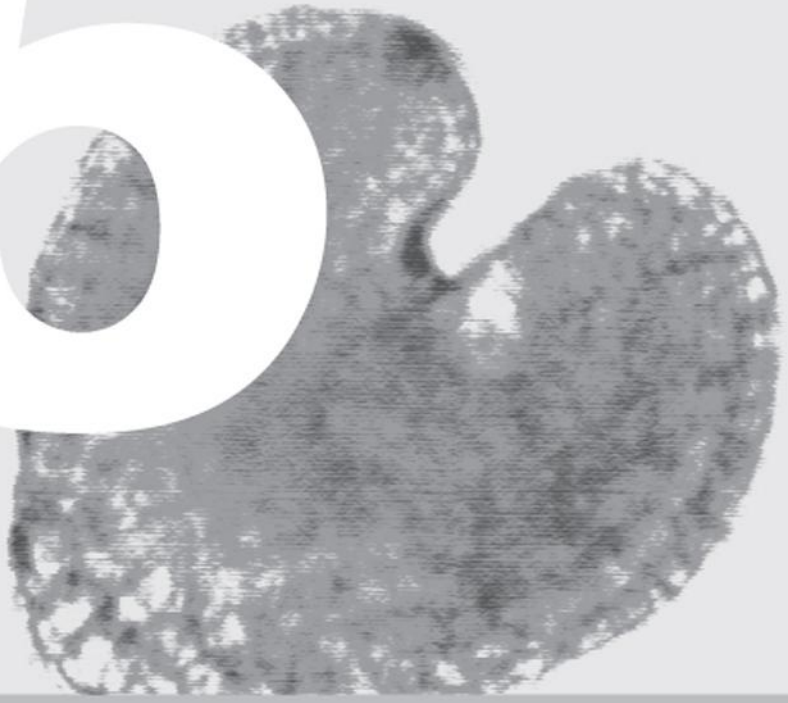
sSNPs in embryogenic cell lines remains unclear and requires additional research in order to clarify the observation.

Regrettably, the individual analysis performed for each *AOX* ORF and intron 1 sequence did not allow the establishment of a direct correction amongst the existence of any regulatory element at a specific polymorphic position and the development of a specific SE efficiency phenotype.

Despite of this, IMEter results points *AOX2a* and *AOX2b* as the genes with improved capacities to increase gene activities, being *AOX2b* the one where the biggest changes amongst cell lines were reported. In the same way, also the intron 1 *AOX2b* IMEter results seem to reinforce the observation that *AOX2b* could be associated with embryogenic capacity, with the detection of higher scores associated mainly with embryogenic cell lines sequences.

Nonetheless, *AOX2b* had been less studied in *Daucus* and the observations presented here may represent a switch on that issue, reinforcing the need for further research on this gene, in order to achieve improved conclusions considering all gene regions at the system, species and individuals level as referred by Nogales *et al.* (2015).

5



CONCLUSIONS

5 - Conclusions

This thesis aimed to contribute for a broad understanding and in-depth characterization of *Daucus* accessions concerning SE induction and efficient expression. It further aimed to provide basic genomic data required for exploring *Daucus AOX* polymorphic regions, helping going forward the research on functional marker development based on cell reprogramming events, in the view of plant breeding for stress tolerance. As highlights, the present study contributed especially through:

- The development of an innovative SE phenotyping approach, using a two stage selection methodology based on growth dynamics evaluation and imaging technology;

- The characterization of 25 *Daucus* accessions concerning SE efficiency responses, using a 139 set of cell lines, which allowed the development and establishment of a basic collection of eight cell lines with stable differential SE efficiencies over time and replicates;

- The amplification of 290 *AOX* sequences from the established collection of cell lines, leading to the identification of 47 unique sequences, comprising the three *AOX* genes identified on the species;

- The characterization of the detected polymorphic positions across the 47 sequences, with the identification of new variants at the size and sequence level associated with specific *Daucus* accessions;

- The indication of polymorphic *AOX2b* sequences in *Daucus*, as possibly correlated with the capacity to develop embryos.

These outcomes strengthened the need for a complete amplification and analysis of *AOX* genes, including UTRs, promoters and additional up and down stream regions, which may be relevant for the regulation of gene expression in the studied species and biological process. In order to achieve comprehensive results, equally important

would be the development of a coherent protocol for polymorphism search and analysis, at each specific region. Although additional work may be required to achieve a full assessment and to establish the final association, acquired data seems to support the original hypothesis, that *AOX* polymorphisms, especially the ones found at the *AOX2b* level, can be associated with SE expression efficiency in *Daucus*. Attained results also provided a new set of phenogenomics data, which added new grounds at the AOX research in *Daucus*, impelling it to go a step forward in the understanding of the polymorphisms occurrence.

BIBLIOGRAPHIC REFERENCES

6 - Bibliographic references

- Abe, F., Saito, K., Miura, K. and Toriyama, K. (2002). A single nucleotide polymorphism in the alternative oxidase gene among rice varieties differing in low temperature tolerance. *FEBS Letters*, 527(1-3):181-185.

- Afuape, S. O., Sayre, R., Tawanda, Z. and Kahya, S. S. (2013). Transgenic cassava lines carrying heterologous alternative oxidase (AtAOX1a) showed impaired quantitative and qualitative response to embryogenesis. *African Journal of Biotechnology*, 12(27):4303-4309.

- Agarwal, M., Shrivastava, N. and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 27(4):617-631.

- Ahn, M. J., Noh, S. A., Ha, S. H., Back, K., Lee, S. W. and Bae, J. M. (2012). Production of ketocarotenoids in transgenic carrot plants with an enhanced level of β-carotene. *Plant Biotechnology Reports*, 6(2):133-140.

- Albury, M. S., Elliott, C. and Moore, A. L. (2009). Towards a structural elucidation of the alternative oxidase in plants*. Physiologia Plantarum*, 137(4):316-327.

- Allard, R. W. (1960). *Principles of plant breeding*. New York: Wiley.

- Amirsadeghi, S., Robson, C. A. and Vanlerberghe, G. C. (2007). The role of the mitochondrion in plant responses to biotic stress. *Physiologia Plantarum*, 129:253:266.

- Ammirato, P. V. (1986). Carrot. In D. A. Evans, W. R. Sharp and P. V. Ammirato (Eds.), *Handbook of plant cell culture*. (pp. 457-499). New York: Macmillan.

- Ammirato, P. V. (1987). Organizational events during somatic embryogenesis. In C. E. Green, D. A. Somers, W. P. Hackett and D. D. Biesboer (Eds.), *Plant tissue and cell culture*. (pp. 57-81). New York: Alan Liss Inc.

- Andersen, J. R. and Lübberstedt, T. (2003). Functional markers in plants. *Trends in Plant Science*, 8(11):554-560.

- Andersson, M. E. and Nordlund, P. (1999). A revised model of the active site of alternative oxidase. *FEBS Letters*, 449(1):17-22.

- Arnholdt-Schmitt, B. (2004). Stress-induced cell reprogramming. A role for global genome regulation?. *Plant Physiology*, 136(1):2579-2586.

- Arnholdt-Schmitt, B. (2005). Functional markers and a 'systemic strategy': convergency between plant breeding, plant nutrition and molecular biology. *Plant Physiology and Biochemistry*, 43(9):817-820.

- Arnholdt-Schmitt, B. (2009). Alternative oxidase (AOX) and stress tolerance-approaching a scientific hypothesis. *Physiologia Plantarum*, 137:314-315.

- Arnholdt-Schmitt, B., Costa, J. H. and Fernandes de Melo, D. (2006). AOX - A functional marker for efficient cell reprogramming under stress?. *Trends in Plant Science*, 11(6):281-287.

- Arumuganathan, K. and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9:208-218.

- Aslam, J., Mujib, A. and Sharma, M. P. (2011). Influence of freezing and non-freezing temperature on somatic embryogenesis and vinblastine production in *Catharanthus roseus* (L.) G. Don. *Acta Physiologiae Plantarum*, 33:473-480.

- Banga, O. (1957a). Origin of the European cultivated carrot. *Euphytica*, 6:54-63.

- Banga, O. (1957b). The development of the original European carrot material. *Euphytica*, 6:64-76.

- Banga, O. (1963). *Main types of the Western carotene carrot and their origin*. (pp. 153). W. E. J. Tjeenk Willink, Zwolle, The Netherlands.

- Baranski, R., Maksylewicz-Kaul, A., Nothnagel, T., Cavagnaro, P. F., Simon, P. W. and Grzebelus, D. (2012). Genetic diversity of carrot cultivars revealed by analysis of SSR loci. *Genetic Resources and Crop Evolution*, 59:163-170.

- Bardini, M., Lee, D., Donini, P., Mariani, A., Gianì, S., Toschi, M., Lowe, C. and Breviario, D. (2004). Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. *Genome*, 47(2):281-291.

- Bartoli, C. G., Gómez, F., Gergoff, G., Guiamet, J. J. and Puntarulo, S. (2005). Up-regulation of the mitochondrial alternative oxidase pathway enhances photosynthetic electron transport under drought conditions. *Journal of Experimental Botany*, 56:1269-1276.

- Bassuner, B. M. R., Lam, R., Lukowitz, W. and Yeung, E. C. (2007). Auxin and root initiation in somatic embryos of *Arabidopsis*. *Plant Cell Reports*, 26:1-11.

- Baurain, D., Dinant, M., Coosemans, N. and Matagne, R. F. (2003). Regulation of the alternative oxidase Aox1 Gene in *Chlamydomonas reinhardtii*. Role of the nitrogen source on the expression of a reporter gene under the control of the Aox1 promoter. *Plant Physiology*, 131:1418-1430.

- Bayliss, M. W. (1976). Variation of cell cycle duration within suspension cultures of *Daucus carota* and its consequence for the induction of ploidy changes with colchicine. *Protoplasma*, 88:279-285.

- Bendall, D. S. and Bonner, W. D. (1971). Cyanide-insensitive respiration in plant mitochondria. *Plant Physiology*, 47:263-245.

- Bennett, M. D. (2004). Perspectives on polyploidy in plants - ancient and neo. *Biological Journal of the Linnean Society*, 82(4):411-423.

- Bennett, M. D. and Leitch, I. J. (1995). Nuclear DNA amounts in angiosperms. *Annals of Botany*, 76:113-176.

- Bennett, M. D. and Smith, J. B. (1976). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society B*, 274:227-274.

- Berthold, D. A. (1998). Isolation of mutants of the *Arabidopsis thaliana* alternative oxidase (ubiquinol: oxygen oxidoreductase) resistant to salicylhydroxamic acid. *Biochimica et Biophysica Acta-Bioenergetics*, 1364(1):73-83.

- Berthold, D. A. and Stenmark, P. (2003). Membrane-bound diiron carboxylate proteins. *Annual Review of Plant Biology*, 54:497-517.

- Berthold, D. A., Andersson, M. E. and Nordlund, P. (2000). New insight into the structure and function of the alternative oxidase. *Biochimica et Biophysica Acta-Bioenergetics*, 1460(2):241-254.

- Bimboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6):1513-1523.

- Bonga, J. M., Klimaszewska, K. and Von Aderakas, P. (2010). Recalcitrance in clonal propagation, in particular of conifers. *Plant Cell, Tissue and Organ Culture*, 100:241-254.

- Borodovsky, M. and McIninch, J. (1993). GENMARK: parallel gene recognition for both DNA strands. *Computers & Chemistry*, 17(2):123-133.

- Bradeen, J. M., Bach, I. C., Briard, M., le Clerc, V., Grzebelus, D., Senalik, D. A. and Simon, P. W. (2002). Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. *Journal of the American Society for Horticultural Science*, 127:383-391.

- Braglia, L., Manca, A., Mastromauro, F. and Breviario, D. (2010). cTBP: a successful intron length polymorphism (ILP)-based genotyping method targeted to well defined experimental needs. *Diversity*, 2(4):572-585.

- Brambilla, V., Battaglia, R., Colombo, M., Masiero, S., Bencivenga, S., Kater, M. M. and Colombo, L. (2007). Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in *Arabidopsis*. *The Plant Cell*, 19(8):2544-2556.

- Breviario, D., Baird, W. V., Sangoi, S., Hilu, K., Blumetti, P. and Gianì, S. (2007). High polymorphism and resolution in targeted fingerprinting with combined β-tubulin introns. *Molecular Breeding*, 20(3):249-259.

- Breviario, D., Gianì, S., Ponzoni, E., Mastromauro, F. and Morello, L. (2008). Plant tubulin intronics. *Cell Biology International*, 32(5):571-573.

- Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J. and Huang, E. S. (2004). Are protein-protein interfaces more conserved in sequence than the rest of the protein surface?. *Protein Science*, 13(1):190-202.

- Campos, M. D., Cardoso, H. G., Linke, B., Costa, J. H., De Melo, D. F., Justo, L., Frederico, A. M. and Arnholdt-Schmitt, B. (2009). Differential expression and co-regulation of carrot AOX genes (*Daucus carota*). *Physiologia Plantarum*, 137(4):578-591.

- Campos, M. D., Frederico, A. M., Nothnagel, T., Arnholdt-Schmitt, B. and Cardoso, H. (2015). Selection of suitable reference genes for reverse transcription quantitative real-time PCR studies on different experimental systems from carrot (*Daucus carota L.*). *Scientia Horticulturae*, 186:115-123.

- Campos, M. D., Nogales, A., Cardoso, H. G., Kumar, S. R., Nobre, T., Sathishkumar, R. and Arnholdt-Schmitt, B. (2016). Stress-induced accumulation of *DcAOX1* and *DcAOX2a* transcripts coincides with critical

time point for structural biomass prediction in carrot primary cultures (*Daucus carota* L.). *Frontiers in Genetics*, 7.

- Capra, J. A. and Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics,* 23(15):1875-1882.

- Capron, A., Chatfield, S., Provart, N. and Berleth, T. (2009). Embryogenesis: pattern formation from a single cell. *The arabidopsis book/American Society of Plant Biologists*, 7:e0126.

- Cardoso, H. G. and Arnholdt-Schmitt, B. (2013). Functional marker development across species in selected traits. In T. Lübberstedt and R. K. Varshney (Eds.), *Diagnostics in Plant Breeding*. (pp. 467-515). Netherlands: Springer.

- Cardoso, H. G., Campos, M. D., Costa, A. R., Campos, M. C., Nothnagel, T. and Arnholdt-Schmitt, B. (2009). Carrot alternative oxidase gene AOX2a demonstrates allelic and genotypic polymorphisms in intron 3. *Physiologia Plantarum*, 137:592-608.

- Cardoso, H. G., Nogales, A., Frederico, A. M., Svensson, J. T., Macedo, E. S., Valadas, V. and Arnholdt-Schmitt, B. (2015). Exploring AOX gene diversity. In K. J. Gupta, L. A. J. Mur and B. Neelwarne (Eds.), *Alternative Respiratory Pathways in Higher Plants*. (pp. 239-254). Chichester: John Wiley & Sons.

- Cardoso, H., Campos, M. D., Nothnagel, T. and Arnholdt-Schmitt, B. (2011). Polymorphisms in intron 1 of carrot AOX2b - A useful tool to develop a functional marker?. *Plant Genetic Resources*, 9(02):177-180.

- Chen, W. P. and Punja. Z. K. (2002). Transgenic herbicide- and disease-tolerant carrot (*Daucus carota* L.) plants obtained through *Agrobacterium*-mediated transformation. *Plant Cell Reports*, 20:929-935.

- Chibbar, R. N., Shyluk, J., Georges, F., Mallard, C. S. and Constabel, F. (1988). Esterase isozymes as markers of somatic embryogenesis in cultured carrot cells. *Journal of Plant Physiology*, 133:367-370.

- Chocobar-Ponce, S., Prado, C., González, R., Rosa, M. and Prado, F. E. (2014). Mitochondrial respiration: involvement of the alternative respiratory pathway and residual respiration in abiotic stress responses. In P. Ahmad and M. R. Wani (Eds.), *Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment.* (pp. 321-364). New York: Springer.

- Chorev, M. and Carmel, L. (2012). The function of introns. *Frontiers in Genetics*, 3(55):1-15.

- Chugh, A. and Eudes, F. (2007). Translocation and nuclear accumulation of monomer and dimmer of HIV-1 Tat basic domain in triticale mesophyll protoplasts. *Biochimica et Biophysica Acta-Biomembranes*, 1768:419-426.

- Chugh, A. and Khurana, P. J. (2002). Gene expression during somatic embryogensis - recent advances. *Current Science*, 83:715-730.

- Clifton, R., Lister, R., Parker, K. L., Sappl, P. G., Elhafez, D., Millar, A. H., Day, D. A. and Whelan, J. (2005). Stress-induced co-expression of alternative respiratory chain components in *Arabidopsis thaliana*. *Plant Molecular Biology*, 58:193-212.

- Clifton, R., Millar, A. H. and Whelan, J. (2006). Alternative oxidases in *Arabidopsis*: a comparative analysis of differential expression in the gene family provides new insights into function of non-phosphorylating bypasses. *Biochimica et Biophysica Acta*, 1757:730-741.

- Coca, M. A., Almoguera, C. and Jordano, J. (1994). Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and functional implications. *Plant Molecular Biology*, 25:479-492.

- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142:169-196.

- Considine M. J., Holtzapffel, R. C., Day. D. A., Whelan, J. and Millar, A. H. (2002). Molecular distinction between alternative oxidase from monocots and dicots. *Plant Physiology*, 129:949-953.

- Costa, J. H., Cardoso, H. G., Campos, M. D., Zavattieri, A., Frederico, A. M., Fernandes de Melo, D. and Arnholdt-Schmitt, B. (2009). *Daucus carota* L. - An old model for cell reprogramming gains new importance through a novel expansion pattern of alternative oxidase (AOX) genes. *Plant Physiology and Biochemistry*, 47:753-759.

- Costa, J. H., Cardoso, H. G., Campos, M. D., Zavattieri, A., Frederico, A. M., Fernandes de Melo, D. and Arnholdt-Schmitt, B. (2014b). Corrigendum to "*Daucus carota* L. - An old model for cell reprogramming gains new importance through a novel expansion pattern of alternative oxidase (AOX) genes" [Plant Physiol. Biochem. 47 (2009) 753-759]". *Plant Physiology and Biochemistry*, 85:114.

- Costa, J. H., Jolivet, Y., Hasenfratz-Sauder, M. P., Orellano, E. G., Da Guia Silva Lima, M., Dizengremel, P. and Fernandes de Melo, D. (2007). Alternative oxidase regulation in roots of *Vigna unguiculata* cultivars differing in drought/salt tolerance. *Journal of Plant Physiology*, 164:718-727.

- Costa, J. H., McDonald, A. E., Arnholdt-Schmitt, B. and Fernandes de Melo, D. (2014a). A classification scheme for alternative oxidases reveals the taxonomic distribution and evolutionary history of the enzyme in angiosperms. *Mitochondrion*, 19:172-183.

- Coutos-Thevenot, P., Jouanneau, J. P., Brown, S., Petiard, V. and Guern, J. (1990). Embryogenic and non-embryogenic cell lines of *Daucus carota* cloned from

meristematic cell clusters: relation with cell ploidy determined by flow cytometry. *Plant Cell Reports*, 8:605-608.

- Crichton, P. G., Affourtit, C., Albury, M. S., Carré, J. E. and Moore, A. L. (2005). Constitutive activity of *Sauromatum guttatum* alternative oxidase in *Schizosaccharomyces pombe* implicates residues in addition to conserved cysteines in α-keto acid activation. *FEBS Letters*, 579(2):331-336.

- Crichton, P. G., Albury, M. S., Affourtit, C. and Moore, A. L. (2010). Mutagenesis of the *Sauromatum guttatum* alternative oxidase reveals features important for oxygen binding and catalysis. *Biochimica et Biophysica Acta-Bioenergetics*, 1797(6):732-737.

- De Jong, A. J., Heidstra, R., Spaink, H. P., Hartog, M. V., Meijer, E. A., Hendriks, T., Schiavo, F. L., Terzi, M., Bisseling, T., Van Kammen, A. and De Vries, S. C. (1993). *Rhizobium* lipooligosaccharides rescue a carrot somatic embryo mutant. *Plant Cell*, 5(6):615-620.

- De-la-Peña, C., Nic-Can, G. I., Galaz-Ávalos, R. M., Avilez-Montalvo, R. and Loyola-Vargas, V. M. (2015). The role of chromatin modifications in somatic embryogenesis in plants. *Frontiers in Plant Science*, 6:635.

- Delporte, F., Jacquemin, J.-M., Masson, P. and Watillon, B. (2012). Insights into the regenerative property of plant cells and their receptivity to transgenesis: wheat as a research case study. *Plant Signaling & Behavior*, 7:1608-1620.

- Delporte, F., Muhovski, Y., Pretová, A. and Watillon, B. (2013). Analysis of expression profiles of selected genes associated with the regenerative property and the receptivity to gene transfer during somatic embryogenesis in *Triticum aestivum* L. *Molecular Biology Reports*, 40(10):5883-5906.

- Deo, P. C., Taylor, M., Harding, R. M., Tyagi, A. P. and Becker, D. K. (2010). Initiation of embryogenic cell suspensions of taro (*Colocasia esculenta* var.

esculenta) and plant regeneration. *Plant Cell, Tissue and Organ Culture*, 100:283-291.

- Djajanegara, I., Finnegan, P. M., Mathieu, C., McCabe, T., Whelan, J. and Day, D. A. (2002). Regulation of alternative oxidase gene expression in soybean. *Plant Molecular Biology*, 50:735-742.

- Dodeman, V. L., Ducreux, G. and Kreis, M. (1997). Zygotic embryogenesis versus somatic embryogenesis. *Journal of Experimental Botany*, 48:1493-1509.

- Dudits, D., Bögre, L. and Györgyey, J. (1991). Molecular and cellular approaches to the analysis of plant embryo development from somatic cells *in vitro*. *Journal of Cell Science*, 99:475-484.

- Dyachok, J. V., Tobin, A. E., Price, N. P. J. and Von Arnold, S. (2000). Rhizobial Nod factors stimulate somatic embryo development in *Picea abies*. *Plant Cell Reports*, 19:290-297.

- Ederli, L., Morettini, R., Borgogni, A., Wasternack, C., Miersch, O., Reale, L., Ferranti, F., Tosti, N. and Pasqualini, S. (2006). Interaction between nitric oxide and ethylene in the induction of alternative oxidase in ozone-treated tobacco plants. *Plant Physiology*, 142:595-608.

- Edwards, D., Forster, J. W., Chagné, D. and Batley, J. (2007). What are SNPs?. In N. C. Oraguzie, E. H. A. Rikkerink, S. E. Gardiner and H. N. Silva (Eds.), *Association mapping in plants*. (pp. 41-52). New York: Springer.

- Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. *SIAM Review*, 21(4):460-480.

- Elhiti, M., Stasolla, C. and Wang, A. M. (2013). Molecular regulation of plant somatic embryogenesis. *In Vitro Cellular & Developmental Biology-Plant*, 49(6):631-642.

- Elmeer, K. E. S. (2013). Factors regulating somatic embryogenesis in plants. In J. Aslam, P. S. Srivastava and M. P. Sharma (Eds.), *Somatic Embryogenesis and Gene Expression*. (pp. 56-81). New Delhi: Narosa Publishing House.

- Emanuelsson, O., Nielsen, H., Brunak, S. and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 300(4):1005-1016.

- Endemann, M., Hristoforoglu, K., Stauber, T. and Wilhelm, E. (2001). Assessment of age-related polyploidy in *Quercus robur* L. somatic embryos and regenerated plants using DNA flow cytometry. *Biologia Plantarum*, 44(3):339-345.

- Eudes, F., Acharya, S., Laroche, A., Selinger, L. B. and Cheng, K.-J. (2003). A novel method to induce direct somatic embryogenesis, secondary embryogenesis and regeneration of fertile green cereal plants. *Plant Cell, Tissue and Organ Culture*, 73:147-157.

- FAOSTAT. (2011). The statistics division of FAO. Available at http://faostat.fao.org/. Accessed 28 Jan 2016.

- Feenstra, K. A., Pirovano, W., Krab, K. and Heringa, J. (2007). Sequence harmony: detecting functional specificity from alignments. *Nucleic Acids Research*, 35(2):495-498.

- Fehér, A. (2008). The initiation phase of somatic embryogenesis: what we know and what we don't. *Acta Biologica Szegediensis*, 52:53-56.

- Fehér, A., Pasternak, T. P. and Dudits, D. (2002). Activation of embyogenic cell division in leaf protoplast-derived alfalfa cells; the role of auxin and stress. *Acta Biologica Szegediensis*, 46:13-14.

- Fehér, A., Pasternak, T. P. and Dudits, D. (2003). Transition of somatic plant cells to an embryogenic state. *Plant Cell, Tissue and Organ Culture*, 74:201-228.

- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 783-791.

- Feng, H., Guan, D., Sun, K., Wang, Y., Zhang, T. and Wang, R. (2013). Expression and signal regulation of the alternative oxidase genes under abiotic stresses. *Acta Biochimica et Biophysica Sinica*, 45(12):985-994.

- Feng, H., Li, X., Duan, J., Li, H. and Liang, H. (2008). Chilling tolerance of wheat seedlings is related to an enhanced alternative respiratory pathway. *Crop Science*, 48:2381-2388.

- Fernie, A. R., Carrari, F. and Sweetlove, L. J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology*, 7:254-261.

- Ferreira, A. O., Cardoso, H. G., Macedo, E. S., Breviario, D. and Arnholdt-Schmitt, B. (2009). Intron polymorphism pattern in AOX1b of wild St John's wort (*Hypericum perforatum*) allows discrimination between individual plants. *Physiologia Plantarum*, 137(4):520-531.

- Finkelstein, R. R., Gampala, S. S. and Rock, C. D. (2002). Abscisic acid signaling in seeds and seedlings. *The Plant Cell*, 14(1):S15-S45.

- Finnegan, P. M., Soole, K. L. and Umbach, A. L. (2004). Alternative mitochondrial electron transport proteins in higher plants. In D. A. Day, A. H. Millar and J. Whelan (Eds.), *Plant mitochondria: From genome to function, Advances in Photosynthesis and Respiration*. Vol. 17, (pp. 163-230). Dordrecht: Kluwer Academic Press.

- Fiorani, F. and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, 64:267-291.

- Fiorani, F., Umbach, A. L. and Siedow, J. N. (2005). The alternative oxidase of plant mitochondria is involved in the acclimation of shoot growth at low

temperature: a study of *Arabidopsis* AOX1a transgenic plants. *Plant Physiology*, 139:1795-1805.

- Fowler, M. R., Ong, L. M., Russinova, E., Atanassov, A. I., Scott, N. W., Slater, A. and Elliott, M. C. (1998). Early changes in gene expression during direct somatic embryogenesis in alfalfa revealed by RAP-PCR. *Journal of Experimental Botany*, 49:249-253.

- Frederico, A. M., Campos, M. D., Cardoso, H. G., Imani, J. and Arnholdt-Schmitt, B. (2009a). Alternative oxidase involvement in *Daucus carota* somatic embryogenesis. *Physiologia Plantarum*, 137:498-508.

- Frederico, A. M., Zavattieri, M. A., Campos, M. D., Cardoso, H. G., McDonald, A. E. and Arnholdt-Schmitt, B. (2009b). The gymnosperm *Pinus pinea* contains both AOX gene subfamilies, AOX1 and AOX2. *Physiologia Plantarum*, 137(4):566-577.

- Fu, X. P., Yang, S. H. and Bao, M. Z. (2008). Factors affecting somatic embryogenesis in anther cultures of Chinese pink (*Dianthus chinensis* L.). *In Vitro Cellular & Developmental Biology-Plant*, 44(3):194-202.

- Fujimura, T. (2014). Carrot somatic embryogenesis. A dream come true?. *Plant Biotechnology Reports*, 8(1):23-28.

- Fung, R. W. M., Wang, C. Y., Smith, D. L., Gross, K. C., Tao, Y. and Tian, M. (2006). Characterization of alternative oxidase (AOX) gene expression in response to methyl salicylate and methyl jasmonate pre-treatment and low temperature in tomatoes. *Journal of Plant Physiology*, 163:1049-1060.

- Gaj, M. D. (2004). Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to *Arabidopsis thaliana* (L.) Heynh. *Plant Growth Regulation*, 43:27-47.

- Galasso, I., Manca, A., Braglia, L., Martinelli, T., Morello, L. and Breviario, D. (2011). h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β-tubulin gene family in *Camelina sativa* (L.) Crantz. *Molecular Breeding*, 28(4):635-645.

- Gallegos, J. E. and Rose, A. B. (2015). The enduring mystery of intron-mediated enhancement. *Plant Science*, 237:8-15.

- Gamborg, O. L., Miller, R. and Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Experimental Cell Research*, 50(1):151-158.

- George, E. F., Hall, M. A. and De Klerk, G.-J. (2008). *Plant propagation by tissue culture*. Chapter 9 (pp. 335-354). Dordrecht: Springer.

- Germanà, M. A., Benedetta, C., Calogero, I. and Rosario, M. (2005). The effect of light quality on anther culture of *Citrus clementina* Hort. ex Tan. *Acta Physiologiae Plantarum*, 27(4):717-721.

- Gilbert, W. (1978). Why genes in pieces?. *Nature*, 271:501.

- Gilbert, W. (1985). Genes-in-pieces revisited. *Science*, 228:823-824.

- Giraud, E., Ho, L. H., Clifton, R., Carroll, A., Estavillo, G., Tan, Y. F. and Whelan, J. (2008). The absence of alternative oxidase 1a in *Arabidopsis* results in acute sensitivity to combined light and drought stress. *Plant Physiology*, 147(2):595-610.

- Giraud, E., Van Aken, O., Ho. L. H. M. and Whelan, J. (2009). The transcription factor ABI4 is a regulator of mitochondrial retrograde expression of alternative oxidase 1a. *Plant Physiology*, 150:1286-1296.

- Gorlova, O., Fedorov, A., Logothetis, C., Amos, C. and Gorlov, I. (2014). Genes with a large intronic burden show greater evolutionary conservation on the protein level. *BMC Evolutionary Biology*, 14(1):1.

- Grafi, G. (2004). How cells dedifferentiate: a lesson from plants. *Developmental Biology*, 268:1-6.

- Grafi, G., Chalifa-Caspi, V., Nagar, T., Plaschkes, I., Barak, S. and Ransbotyn, V. (2011). Plant response to stress meets dedifferentiation. *Planta*, 233:433-438.

- Grant, N., Onda, Y., Kakizaki, Y., Ito, K., Watling, J. and Robinson, S. (2009). Two Cys or not two Cys? That is the question; alternative oxidase in the thermogenic plant sacred lotus. *Plant Physiology*, 150:987-995.

- Gray, G. R., Maxwell, D. P., Villarimo, A. R. and McIntosh, L. (2004). Mitochondria/nuclear signaling of alternative oxidase gene expression occurs through distinct pathways involving organic acids and reactive oxygen species. *Plant Cell Reports*, 23:497-503.

- Grieb, B., Schafer, F., Imani, J., Mashayekhi, K. N., Arnholdt-Schmitt, B. and Neumann, K. H. (1997). Changes in soluble proteins and phytohormone concentrations of cultured carrot petiole explants during induction of somatic embryogenesis (*Daucus carota* L.). *Journal of Applied Botany and Food Quality*, 71:94-103.

- Grzebelus, D., Baranski, R., Spalik, K., Allender, C. and Simon, P. (2011). *Daucus*. In C. Kole (Ed.), *Wild crop relatives: genomic and breeding resources, vegetables*. (pp. 91-113). Berlin: Spring-Verlag.

- Grzebelus, D., Iorizzo, M., Senalik, D., Ellison, S., Cavagnaro, P., Macko-Podgorni, A., Heller-Uszynska, K., Kilian, A., Nothnagel, T., Simon, P. W. and Baranski, R. (2014). Diversity, genetic mapping, and signatures of domestication in the

carrot (*Daucus carota* L.) genome, as revealed by Diversity Arrays Technology (DArT) markers. *Molecular Breeding*, 33:625-637.

- Guharoy, M. and Chakrabarti, P. (2005). Conservation and relative importance of residues across protein-protein interfaces. *Proceedings of the National Academy of Sciences*, 102(43):15447-15452.

- Gupta, P. K. and Rustgi, S. (2004). Molecular markers from the transcribed / expressed region of the genome in higher plants. *Functional & Integrative Genomics*, 4(3):139-162.

- Gutiérrez-Mora, A., González-Gutiérrez, A. G., Rodríguez-Garay, B., Ascencio-Cabral, A. and Li-Wei, L. (2012). Plant somatic embryogenesis: some useful considerations. In K. I. Sato (Ed.), *Embryogenesis*. (pp. 229-241). Rijeka: InTech.

- Haberlandt, G. (1902). Culturversuche mit isolierten Pflanzenzellen. *Sitzungsber Akad Wissensch Wien Math-Naturw K1*, (111):69-92.

- Haga, N., Kobayashi, K., Suzuki, T., Maeo, K., Kubo, M., Ohtani, M. and Ito, M. (2011). Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in *Arabidopsis*. *Plant Physiology*, 157(2):706-717.

- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41:95-98.

- Hampl, V., Pavlícek, A. and Flegr, J. (2001). Construction and bootstrap analysis of DNA fingerprinting-based phylogenetic trees with the freeware program FreeTree: application to trichomonad parasites. *International Journal of Systematic and Evolutionary Microbiology*, 51(3):731-735.

- Han, G. Y., Wang, X. F., Zhang, G. Y. and Ma, Z. Y. (2009). Somatic embryogenesis and plant regeneration of recalcitrant cottons (*Gossypium hirsutum*). *African Journal of Biotechnology*, 8(3).

- Han, Y., Jin, X. L., Wu, F. B. and Zhang, G. P. (2011). Genotypic differences in *callus* induction and plant regeneration from mature embryos of barley (*Hordeum vulgare* L.). *Journal of Zhejiang University Science B*, 12(5):399-407.

- Hauser, T. P. (2002). Frost sensitivity of hybrids between wild and cultivated carrots. *Conservation Genetics*, 3:75-78.

- Hauser, T. P. and Bjorn, G. K. (2001). Hybrids between wild and cultivated carrots in Danish carrot fields. *Genetic Resources and Crop Evolution*, 48:499-506.

- Hauser, T. P., Bjorn, G. K., Magnussen, L. and Shim, S. I. (2004). Hybrids between cultivated and wild carrots: a life history. In H. C. M. den Nijs, D. Bartsch and J. Sweet (Eds.), *Introgression from genetically modified plants into wild relatives*. (pp. 41-51). Wallingford: CABI Publishing.

- Hay, A., Kaur, H., Phillips, A. S., Hedden, P., Hake, S. and Tsiantis, M. (2002). The gibberellin pathway mediates knotted1-type homeobox function in plants with different body plans. *Current Biology*, 12:1557-1565.

- Heazlewood, J. L., Tonti-Filippini, J. S., Gout, A. M., Day, D. A., Whelan, J. and Millar, A. H. (2004). Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *The Plant Cell*, 16(1):241-256.

- Henry, Y., Vain, P. and De Buyser, J. (1994). Genetic analysis of *in vitro* plant tissue culture responses and regeneration capacities. *Euphytica*, 79:45-58.

- Heywood, V. H. (1968). *Daucus* L. In T. G. Tutin, V. H. Heywood, N. A. Burgess, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb (Eds.), *Flora Europaea 2.* (pp. 373-375). Cambridge: Cambridge University Press.

- Heywood, V. H. (1982). Multivariate taxonomic synthesis of the tribe *Caucalideae*. *Monographs in Systematic Botany from the Missouri Botanical Garden*, 6:727-736.

- Ho, L. H., Giraud, E., Lister, R., Thirkettle-Watts, D., Low, J., Clifton, R. and Whelan, J. (2007). Characterization of the regulatory and expression context of an alternative oxidase gene provides insights into cyanide-insensitive respiration during growth and development. *Plant Physiology*, 143(4):1519-1533.

- Holtzapffel, R. C., Castelli, J., Finnegan, P. M., Millar, A. H., Whelan, J. and Day, D. A. (2003). A tomato alternative oxidase protein with altered regulatory properties. *Biochimica et Biophysica Acta-Bioenergetics*, 1606(1):153-162.

- Hussain, M., Fareed, S., Ansari, S., Rahman, M., Ahmad, I. Z. and Mohd, S. (2012). Current approaches toward production of secondary plant metabolites. *Journal of Pharmacy and Bioallied Sciences*, 4:10-20.

- Ikeda, M., Umehara, M. and Kamada, H. (2006). Embryogenesis-related genes: Its expression and roles during somatic and zygotic embryogenesis in carrot and *Arabidopsis*. *Plant Biotechnology*, 23:153-161.

- Ikeda, M., Umehara, M., Satoh, S. and Kamada, H. (2004). Stress-induced somatic embryogenesis in vegetative tissues of *Arabidopsis thaliana*. *The Plant Journal*, 34:107-114.

- Imani, J., Baltruschat, H., Stein, E., Jia, G., Vogelsberg, J., Kogel, K. H. and Hückelhoven, R. (2006). Expression of BAX inhibitor-1 in carrots confers resistance to *Botrytis cinerea*. *Molecular Plant Pathology*, 7:279-284.

- Imani, J., Berting, A., Nitsche, S., Schaefer, S., Gerlich, W. H. and Neumann, K. H. (2002). The integration of a major hepatitis B virus gene into cell-cycle synchronized carrot cell suspension cultures and its expression in regenerated carrot plants. *Plant Cell, Tissue and Organ Culture*, 71:157-164.

- Imani, J., Thi, L. T., Langen, G., Arnholdt-Schmitt, B., Roy, S., Lein, C., Kumar, A. and Neumann, K.-H. (2001). Somatic embryogenesis and DNA organization of genomes from selected *Daucus* species. *Plant Cell Reports*, 20:537-541.

- Iorizzo, M., Senalik, D., Ellison, S., Grzebelus, D., Cavagnaro, P. F., Allender, C., Brunet, J., Spooner, D., Van Deynze, A. and Simon, P. W. (2013). Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (*Apiaceae*). *American Journal of Botany*, 100:930-938.

- Iovene, M., Cavagnaro, P. F., Senalik, D., Buell, C. R., Jiang, J. and Simon P. W. (2011). Comparative FISH mapping of *Daucus* species (*Apiaceae* family). *Chromosome Research*, 19:493-506.

- Iovene, M., Grzebelus, E., Carputo, D., Jiang, J. and Simon, P. W. (2008). Major cytogenetic landmarks and karyotype analysis in *Daucus carota* and other *Apiaceae*. *American Journal of Botany*, 95:793-804.

- IPGRI. (1998). *Descriptors for wild and cultivated carrots* (*Daucus carota* L.). International Plant Genetic Resources Institute, Rome.

- Ishigaki, G., Gondo, T., Rahman, M. M., Umami, N. and Akashi, R. (2014). Spontaneous appearance of polyploids in plants regenerated from embryogenic calli derived from seedling-meristems of ruzigrass (*Brachiaria ruziziensis* Germain et Everard). *Grassland Science*, 60:24-30.

- Jacobsen, H.-J. (1991). Biochemical and molecular studies on plant development in vitro. In B. B. Biswas and J. R. Harris (Eds.), *Subcellular Biochemistry - Plant Genetic Engineering*. (pp. 265-277). New York: Plenum Press.

- Ji, A., Geng, X., Zhang, Y., Yang H. and Wu, G. (2011). Advances in somatic embryogenesis research of horticultural plants. *American Journal of Plant Sciences*, 2:727-732.

- Jia, L., Zhang, B., Mao, C., Li, J., Wu, Y., Wu, P. and Wu, Z. (2008). OsCYT-INV1 for alkaline/neutral invertase is involved in root cell development and reproductivity in rice (*Oryza sativa* L.). *Planta*, 228:51-59.

- Jiang, F., Zhu, J. and Liu, H. L. (2013). Protoplasts: a useful research system for plant cell biology, especially dedifferentiation. *Protoplasma*, 250:1231-1238.

- Jiang, G.-L. (2013). Molecular markers and marker-assisted breeding in plants. In S. B. Andersen (Ed.), *Plant Breeding from Laboratories to Fields*. (pp. 45-84). Rijeka: InTech.

- Jiménez, V. and Thomas, C. (2006). Participation of plant hormones in determination and progression of somatic embryogenesis. In A. Mujib and J. Šamaj (Eds.), *Somatic embryogenesis (Plant Cell Monographs)*. (pp 103-118). Berlin: Springer.

- Jiménez, V. M. (2001). Regulation of *in vitro* somatic embryogenesis with emphasis on to the role of endogenous hormones. *Revista Brasileira de Fisiologia Vegetal*, 13(2):196-223.

- Jin, F., Hu, L., Yuan, D., Xu, J., Gao, W., He, L., Yang, X. and Zhang, X. (2014). Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: evidence for stress response functions in SE development. *Plant Biotechnology Journal*, 12:161-173.

- Jo, B. S. and Choi, S. S. (2015). Introns: The functional benefits of introns in genomes. *Genomics & Informatics*, 13(4):112-118.

- Joshi, R. and Kumar, P. (2013). Regulation of somatic embryogenesis in crops: a review. *Agricultural Review*, 34:1-20.

- Kalinina, O. V., Mironov, A. A., Gelfand, M. S. and Rakhmaninova, A. B. (2004). Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Science*, 13(2):443-456.

- Kamada, H., Ishikawa, K., Saga, H. and Harada, H. (1993). Induction of somatic embryogenesis in carrot by osmotic stress. *Plant Tissue Culture Letters*, 10:38-44.

- Kamada, H., Kobayashi, K., Kiyosue, T. and Harada, H. (1989). Stress induced somatic embryogenesis in carrot and its application to synthetic seed production. *In vitro Cellular and Developmental Biology Plant*, 25(12):1163-1166.

- Kamle, S., Kumar, A. and Bhatnagar, R. K. (2011). Development of multiplex and construct specific PCR assay for detection of cry2Ab transgene in genetically modified crops and product. *GM Crops & Food*, 2:1-8.

- Karami, O. and Saidi, A. (2010). The molecular basis for stress-induced acquisition of somatic embryogenesis. *Molecular Biology Reports*, 37(5):2493-2507.

- Karami, O., Aghavaisi, B. and Pour, A. M. (2009). Molecular aspects of somatic-to-embryogenic transition in plants. *Journal of Chemical Biology*, 2:177-190.

- Karlin, S. and Altschul, S. F. (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. *Proceedings of the National Academy of Sciences*, 90(12):5873-5877.

- Katoh, K., Kuma, K. I., Toh, H. and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, 33(2):511-518.

- Katoh, K., Misawa, K., Kuma, K. I. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059-3066.

- Kikuchi, A., Sanuki, N., Higashi, K., Koshiba, T. and Kamada, H. (2006). Abscisic acid and stress treatment are essential for the acquisition of embryogenic competence by carrot somatic cells. *Planta*, 223:637-645.

- Kitamiya, E., Suzuki, S., Sano, T. and Nagata. T. (2000). Isolation of two genes that were induced upon the initiation of somatic embryogenesis on carrot hypocotyls by high concentrations of 2,4-D. *Plant Cell Reports*,19:551-557.

- Kiyosue, T., Takano, K., Kamada, H. and Harada, H. (1990). Induction of somatic embryogenesis in carrot by heavy metal ions. *Canadian Journal of Botany*, 68(10):2021-2033.

- Komamine, A., Murata, N. and Nomura, K. (2005). Mechanisms of somatic embryogenesis in carrot suspension cultures - morphology, physiology, biochemistry, and molecular biology. *In Vitro Cellular & Developmental Biology-Plant*, 41:6-10.

- Konieczny, R., Sliwinska, E., Pilarska, M. and Tuleja, M. (2012). Morphohistological and flow cytometric analyses of somatic embryogenesis in *Trifolium nigrescens* Viv.. *Plant Cell, Tissue and Organ Culture*, 9:131-141.

- Kooiker M., Airoldi C. A., Losa A., Manzotti P. S., Finzi L., Kater M. M. and Colombo L. (2005). BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic *Arabidopsis* gene SEEDSTICK. *Plant Cell*, 17:722-729.

- Kubaláková, M., Doležel, J. and Lebeda, A. (1996). Ploidy instability of embryogenic cucumber (*Cucumis sativus* L.) *callus* culture. *Biologia Plantarum*, 38(3):475-480.

- Kumar, U. (2000). *Synthetic seeds for commercial crop production*. New Delhi: Agrobios.

- Kurczyńska, E. U., Potocka, I., Dobrowolska, I., Kulinska-Lukaszek, K., Sala, K. and Wrobel, J. (2012). Cellular markers for somatic embryogenesis. In K.-I. Sato (Ed.), *Embryogenesis*. (pp. 307-332). Rijeka: InTech.

- Lambers, H. (1997). Oxidation of mitochondrial NADH and the synthesis of ATP. In D. T. Dennis, D. H. Turpin, D. D. Lefebrve and D. B Layzell (Eds.), *Plant Metabolism*. (pp. 200-219). Singapore: Longman Singapore Publishers.

- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research*, 33(2):299-302.

- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T.J. and Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947-2948.

- Lee, B. Y. and Downie, S. R. (1999). A molecular phylogeny of *Apiaceae* tribe *Caucalideae* and related taxa: inferences based on ITS sequence data. *Systematic Botany*, 24:461-479.

- Lee, B. Y. and Downie, S. R. (2000). Phylogenetic analysis of cpDNA restriction sites and rps16 intron sequences reveals relationships among *Apiaceae* tribes *Caucalideae*, *Scandiceae* and related taxa. *Plant Systematics and Evolution*, 221:35-60.

- Lee, B. Y., Levin, G. A. and Downie, S. R. (2001). Relationships within the spiny-fruited umbellifers (*Scandiceae* subtribes *Daucinae* and *Torilidinae*) as assessed by phylogenetic analysis of morphological characters. *Systematic Botany*, 26:622-642.

- Lee, C. Y., Kim, Y. K., Kim, Y. S., Suh, S. Y., Lee, S. Y. and Park, S. U. (2009b). Somatic embryogenesis and plant regeneration in *Cnidium officinale* Makino. *Journal of Medicinal Plants Research*, 3:96-100.

- Lee, H. O., Davidson, J. M. and Duronio, R. J. (2009a). Endoreplication: polyploidy with purpose. *Genes & Development*, 23(21):2461-2477.

- Li, H. and Kurata, K. (2005). Static suspension culture of carrot somatic embryogenesis. *Journal of Bioscience and Bioengineering*, 99:300-302.

- Li, Z. T., Dhekney, S. A., Dutt, M., Aman, M. V., Tattersall, J., Kelly, K. T. and Gray, D. J. (2006). Optimizing Agrobacterium-mediated transformation of grapevine. *In Vitro Cellular & Developmental Biology-Plant*, 42: 220-227.

- Liang, S., Zhang, C., Liu, S. and Zhou, Y. (2006). Protein binding site prediction using an empirical scoring function. *Nucleic Acids Research*, 34(13):3698-3707.

- Lin, X., Hwang, G. J. and Zimmerman, J. L. (1996). Isolation and characterization of a diverse set of genes from carrot somatic embryos. *Plant Physiology*, 112:1365-1374.

- Luby, C. H. and Goldman, I. L. (2016). Improving freedom to operate in carrot breeding through the development of eight open source composite populations of carrot (*Daucus carota* L. var. *sativus*). *Sustainability*, 8:479.

- Lynch, M. (2002). Intron evolution as a population-genetic process. *Proceedings of the National Academy of Sciences*, 99(9):6118-6123.

- Macko-Podgorni, A., Nowicka, A., Grzebelus, E., Simon, P. W. and Grzebelus, D. (2013). DcSto: carrot Stowaway-like elements are abundant, diverse, and polymorphic. *Genetica*, 141(4-6):255-267.

- Mahdavi-Darvari, F., Noor, N. M. and Ismanizan, I. (2015). Epigenetic regulation and gene markers as signals of early somatic embryogenesis. *Plant Cell, Tissue and Organ Culture*,120(2):1-16.

- Marcotte, W. R., Russell, S. H. and Quatrano, R. S. (1989). Abscisic acid-responsive sequences from the em gene of wheat. *The Plant Cell*, 1(10):969-976.

- Mariano, A. B., Valente, C., Maurer, J. B. B., Cadena, S. M. S. C., Rocha, M. E. M., de Oliveira, M. B. M. and Carnieri, E. G. S. (2008). Functional characterization of mitochondria isolated from the ancient gymnosperm *Araucaria angustifolia*. *Plant Science*, 175(5):701-705.

- Mascarenhas, D., Mettler, I. J., Pierce, D. A. and Lowe, H. W. (1990). Intron-mediated enhancement of heterologous gene expression in maize. *Plant Molecular Biology*, *15*(6):913-920.

- Maxwell, D., Nickels, R. L. and McIntosh, L. (2002). Evidence of mitochondrial involvement in the transduction of signals required for the induction of genes associated with pathogen attack and senescence. *Plant Journal*, 29:263-279.

- McCabe, P. F., Valentine, T. A., Forsberg, L. S. and Pennell, R. I. (1997). Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *The Plant Cell*, 9:2225-2241.

- McDonald, A. E. (2008). Alternative oxidase: an inter-kingdom perspective on the function and regulation of this broadly distributed ‘cyanide-resistant’ terminal oxidase. *Functional Plant Biology*, 35:535-552.

- McDonald, A. E. and Vanlerberghe, G. C. (2004). Branched mitochondrial electron transport in the Animalia: presence of alternative oxidase in several animal phyla. *IUBMB Life*, 56:333-341.

- McDonald, A. E. and Vanlerberghe, G. C. (2006). Origins, evolutionary history, and taxonomic distribution of alternative oxidase and plastoquinol terminal oxidase. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 1:357-364.

- McDonald, A. E., Vanlerberghe, G. C. and Staples, J. F. (2009). Alternative oxidase in animals: unique characteristics and taxonomic distribution. *Journal of Experimental Botany*, 212: 2627-2634.

- Meneses, A., Flores, D., Muñoz, M., Arrieta, G. and Espinoza, A. M. (2005). Effect of 2,4-D, hydric stress and light on indica rice (*Oryza sativa*) somatic embryogenesis. *Revista de Biología Tropical*, 53:361-368.

- Merkle, S. A., Parrot, W. A. and Flinn, B. S. (1995). Morphogenic aspects of somatic embryogenesis. In Thorpe, T. A. (Ed.), *In Vitro Embryogenesis in Plants*. (pp. 155-203). Dordrecht: Kluwer Academic Publishers.

- Mintseris, J. and Weng, Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. *Proceedings of the National Academy of Sciences*, 102(31):10930-10935.

- Mizuno, N., Sugie, A., Kobayashi, F. and Takumi, S. (2007). Mitochondrial alternative pathway is associated with development of freezing tolerance in common wheat. *Journal of Plant Physiology*, 165(4):462-467.

- Moabbi, A. M., Agarwal, N., El Kaderi, B. and Ansari, A. (2012). Role for gene looping in intron-mediated enhancement of transcription. *Proceedings of the National Academy of Sciences*, 109(22):8505-8510.

- Moore, A. L. and Albury, M. S. (2008). Further insights into the structure of the alternative oxidase: from plants to parasites. *Biochemical Society Transactions*, 36(5):1022-1026.

- Moore, A. L., Albury, M. S., Crichton, P. G. and Affourtit, C. (2002). Function of the alternative oxidase: is it still a scavenger? *Trends Plant Science.*, 7:478-481.

- Moore, A. L., Carré, J. E., Affourtit, C., Albury, M. S., Crichton, P. G., Kita, K. and Heathcote, P. (2008). Compelling EPR evidence that the alternative oxidase is a diiron carboxylate protein. *Biochimica et Biophysica Acta-Bioenergetics*, 1777(4):327-330.

- Morita, S., Tsukamoto, S., Sakamoto, A., Makino, H., Nakauji, E., Kaminaka, H. and Tanaka, K. (2012). Differences in intron-mediated enhancement of gene expression by the first intron of cytosolic superoxide dismutase gene from rice in monocot and dicot plants. *Plant Biotechnology*, 29(1):115-119.

- Murakami, Y. and Toriyama K. (2008). Enhanced high temperature tolerance in transgenic rice seedlings with elevated levels of alternative oxidase, OsAOX1a. *Plant Biotechnology*, 25:361-364.

- Murtagh, F. (1984). Complexities of hierarchic clustering algorithms: state of the art. *Computational Statistic Quarterly*, 1(2):101-113.

- Naing, A. H., Kim, C. K., Yun, B. J., Jin, J. Y. and Lim, K. B. (2013). Primary and secondary somatic embryogenesis in *Chrysanthemum* cv. Euro. *Plant Cell, Tissue and Organ Culture*, 112(3):361-368.

- Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H. and Yamaguchi-Shinozaki, K. (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. *The Plant Journal*, 34(2):137-148.

- Nasim, S. A., Mujib, A., Kapoor, R., Fatima, S., Junaid, A. and Mahmooduzzafar (2010). Somatic embryogenesis in *Allium sativum* L. (cv. Yamuna Safed 3): Improving embryo maturation and germination with PGRs and carbohydrates. *Anales de Biología*, 32:1-9.

- Naydenov, N. G., Khanam, S. M., Atanassov, A. and Nakamura, C. (2008). Expression profiles of respiratory components associated with mitochondrial biogenesis during germination and seedling growth under normal and restricted conditions in wheat. *Genes & Genetic Systems*, 83(1):31-41.

- Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10):5269-5273.

- Nic-Can, G. I., Galaz-Ávalos, R. M., De-la-Peña, C., Alcazar-Magaña, A., Wrobel, K. and Loyola-Vargas, V. M. (2015). Somatic embryogenesis: Identified factors that lead to embryogenic repression. A case of species of the same genus. *PLoS ONE,* 10(6):0126414.

- Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering, Design and Selection*, 10:1-6.

- Nishiwaki, M., Fujino, K., Koda, Y., Masuda. K. and Kikuta, Y. (2000). Somatic embryogenesis induced by the simple application of abscisic acid to carrot (*Daucus carota* L.) seedlings in culture. *Planta*, 211:756-759.

- Noctor, G., De Paepe, R. and Foyer, C. H. (2007). Mitochondrial redox biology and homeostasis in plants. *Trends Plant Science*, 12:125:134.

- Nogales, A., Noceda, C., Ragonezzi, C., Cardoso, H., Campos, M.D., Frederico, A. M., Sircar, D., Iyer, R., Polidoros, A., Peixe, A. and Arnholdt-Schmit, B. (2015). Functional marker development from AOX genes requires deep phenotyping and individualized diagnosis. In K. J. Gupta, L. A. J. Mur and B. Neelwarne (Eds.), *Alternative Respiratory Pathways in Higher Plants*. (pp. 275-280). Chichester: John Wiley & Sons.

- Nogales, A., Nobre, T., Cardoso, H. G., Muñoz-Sanhueza, L., Valadas, V., Campos, M. D. and Arnholdt-Schmitt, B. (2016). Allelic variation on DcAOX1 gene in

carrot (*Daucus carota* L.): An interesting simple sequence repeat in a highly variable intron. *Plant Gene*, 5:49-55.

- Nomura, k. and Komamine, A. (1995). Physiological and biochemical aspects of somatic embryogenesis. In Thorpe, T. A. (Ed.), *In vitro embryogenesis in plants*. (pp. 249-266). Dordrecht: Kluwer Academic.

- Ochatt, S. J., Atif, R. M., Patat-Ochatt, E. M., Jacas, L. and Conreux, C. (2010). Competence versus recalcitrance for *in vitro* regeneration. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 38:102-108.

- Ordog, S. H., Higgins, V. J. and Vanlerberghe, G. C. (2002). Mitochondrial alternative oxidase is not a critical component of plant viral resistance but may play a role in the hypersensitive response. *Plant Physiology*, 129(4):1858-1865.

- Păcurar, D. I., Păcurar, M. L., Street, N., Bussell, J. D., Pop, T. I., Gutierrez, L. and Bellini, C. (2012). A collection of INDEL markers for map-based cloning in seven *Arabidopsis* accessions. *Journal of Experimental Botany*, 63(7):2491-2501.

- Panchenko, A. R., Kondrashov, F. and Bryant, S. (2004). Prediction of functional sites by analysis of sequence and structure conservation. *Protein Science*, 13(4):884-892.

- Parra, G., Bradnam, K., Rose, A. B. and Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Research*, 39(13):5328-5337.

- Pasternak, T. P., Prinsen, E., Ayadin, F., Miskolczi, P., Potters, G., Asard, H., Vanonckelen, H. A., Dudits, D. and Fehér, A. (2002). The role of auxin, pH, and stress in the activation of embryogenic cell division in leaf protoplast-derived cells of alfalfa. *Plant Physiology*, 129(4):1807-1819.

- Pavlicek, A., Hrda, S. and Flegr, J. (1999). FreeTree - freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of the genus *Frenkelia*. *Folia Biologica*, 45:97-9.

- Pedrosa, L. and Vasil, I. K. (1996). Optimization of somatic embryogenesis and longterm regeneration in *callus* cultures of diploperennial teosinte (*Zea diploperennis* Iltis, Doebley & Guzman). *Maydica*, 41:333-348.

- Pennell, R. I., Cronk, Q. C., Forsberg, L. S., Stöhr, C., Snogerup, L., Kjellbom, P. and McCabe, P. F. (1995). Cell-context signalling. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences Royal Society*, 350(1331):87-93.

- Pennell, R. I., Janniche, L., Scofield, G. N., Booij, H., De Vries, S. C. and Roberts, K. (1992). Identification of a transitional cell state in the developmental pathway to carrot somatic embryogenesis. *The Journal of Cell Biology*, 119:1371-1380.

- Pérez-Clemente, R. A. and Gómez-Cadenas, A. (2012). *In vitro* tissue culture, a tool for the study and breeding of plants subjected to abiotic stress conditions. In A. Leva and L. Rinaldi (Eds.), *Recent Advances in Plant in vitro Culture*. (pp. 91-108). Rijeka: InTech.

- Peters, S., Imani, J., Mahler, V., Foetisch, K., Kaul, S., Paulus, K., Scheurer, S., Vieths, S and. Kogel, K. H. (2011). Dau c 1.01 and Dau c 1.02-silenced transgenic carrot plants show reduced allergenicity to patients with carrot allergy. *Transgenic Research*, 20:547-556.

- Peterson, C. E. and Simon P. W. (1986). Carrot Breeding. In M. J. Bassett (Ed.), *Breeding Vegetable Crops*. (pp. 321-356). Westport: AVI Publishing.

- Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E. and Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 424(6944):85-88.

- Pirovano, W., Feenstra, K. A. and Heringa, J. (2006). Sequence comparison by sequence harmony identifies subtype-specific functional sites. *Nucleic Acids Research*, 34(22):6540-6548.

- Plaxton, W. C. and Podestá, F. E. (2006). The functional organization and control of plant respiration. *Critical Reviews in Plant Sciences*, 25(2):159-198.

- Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J. P. and Hyvönen, J. (2013). Advances in plant gene-targeted and functional markers: a review. *Plant Methods*, 9(1):6.

- Polidoros, A. N., Mylona P. V. and Arnholdt-Schmitt B. (2009). Aox gene structure, transcript variation and expression in plants. *Physiologia Plantarum*, 137:342-353.

- Polidoros, A. N., Mylona, P. V., Pasentsis, K., Scandalios, J. G. and Tsaftaris, A. S. (2005). The maize alternative oxidase 1a (Aox1a) gene is regulated by signals related to oxidative stress. *Redox Report*, 10:71-78.

- Potters, G., Pasternak, T. P., Guisez, Y. and Jansen, M. A. K. (2009). Different stresses, similar morphogenic responses: integrating a plethora of pathways. *Plant, Cell & Environment*, 32:158-169.

- Potters, G., Pasternak, T. P., Guisez, Y., Palme, K. J. and Jansen, M. A. K. (2007). Stress-induced morphogenic responses: growing out of trouble?. *Trends in Plant Science*, 12:98-105.

- Pu, X. J., Lv, X. and Lin, H. H. (2015). Unraveling the evolution and regulation of the alternative oxidase gene family in plants. *Development Genes and Evolution*, 225(6):331-339.

- Punja, Z. K., Jayaraj, J. and Wally, O. (2007). Carrot. In Pua E. C. and Davey M. R. (Eds.), *Biotechnology in Agriculture and Forestry: Transgenic Crops IV*. (pp. 277-294). Berlin: Springer-Verlag.

- Purvis, A. C. and Shewfelt, R. L. (1993). Does the alternative pathway ameliorate chilling injury in sensitive plant tissues?. *Physiologia Plantarum*, 88:712-718.

- Quiroz-Figueroa, F., Rojas-Herrera, R., Galaz-Avalos, R. M. and Loyola-Vargas, V. M. (2006). Embryo production through somatic embryogenesis can be used to study cell differentiation in plants. *Plant Cell, Tissue and Organ Culture*, 86:285-301.

- Raemakers, C. J. J. M., Jacobsen, E. and Visser, R. G. F. (1995). Secondary somatic embryogenesis and applications in plant breeding. *Euphytica*, 81:93-107.

- Raghavan, V. (2004). Role of 2,4-dichlorophenoxyacetic acid (2,4-D) in somatic embryogenesis on cultured zygotic embryos of *Arabidopsis*: cell expansion, cell cycling, and morphogenesis during continuous exposure of embryos to 2,4-D. *American Journal of Botany*, 91:1743-1756.

- Raghavan, V. (2006). Can carrot and *Arabidopsis* serve as model systems to study the molecular biology of somatic embryogenesis?. *Current Science*, 907(1):1336-1343.

- Raghavendra, A. S. and Padmasree, K. (2003). Beneficial interactions of mitochondrial metabolism with photosynthetic carbon assimilation. *Trends Plant Science*, 8:546-553.

- Rahmawati, A. and Esyanti, R. R. (2014). Analysis of secondary metabolite production in somatic embryo of Pasak Bumi (*Eurycoma longifolia* Jack.). *Procedia Chemistry*, 13:112-118.

- Rai, M. K., Kalia, R. K., Singh, R., Gangola, M. P. and Dhawan, A. K. (2011). Developing stress tolerant plants through *in vitro* selection - an overview of the recent progress. *Environmental and Experimental Botany*, 71:89-98.

- Rao, K. S. (1996). Embryogenesis in flowering plants - recent approaches and prospects. *Journal of Biosciences*, 21:827-841.

- Rasmusson, A. G., Fernie, A. R. and Van Dongen, J. T. (2009). Alternative oxidase: a defence against metabolic fluctuations?. *Physiologia Plantarum*,137:371-382.

- Reddy, M. C., Murthy, K. S. R. and Pullaiah, T. (2012). Synthetic seeds: a review in agriculture and forestry. *African Journal of Biotechnology*, 11:14254-14275.

- Reinert, J. (1958). Über die kontrolle der morphogenese und die induktion von adventivembryonene an gewebekulturen aus karotten. *Planta*, 53:318-333.

- Rhoads, D. M. and Subbaiah, C. C. (2007). Mitochondrial retrograde regulation in plants. *Mitochondrion*, 7:177-194.

- Rhoads, D. M., Umbach, A. L., Sweet, C. R., Lennon, A. M., Rauch, G. S. and Siedow, J. N. (1998). Regulation of the cyanide-resistant alternative oxidase of plant mitochondria: identification of the cysteine residue involved in α-keto acid stimulation and inter subunit disulfide bond formation. *Journal of Chemical Biology*, 273:30750-30756.

- Ronchi, V. N., Giorgetti, L., Tonelli, M. and Martini, G. (1992). Ploidy reduction and genome segregation in cultured carrot cell lines. II. Somatic meiosis. *Plant Cell, Tissue and Organ Culture*, 30(2):115-119.

- Rong, J., Janson, S., Umehara, M., Ono, M. and Vrieling, K. (2010). Historical and contemporary gene dispersal in wild carrot (*Daucus carota* ssp. *carota*) populations. *Annals of Botany*, 106:285-296.

- Rosales-Mendoza, S. and Tello-Olea, M. A. (2015). Carrot cells: a pioneering platform for biopharmaceuticals production. *Molecular Biotechnology*, 57:219-232.

- Rose, A. B. (2008). Intron-mediated regulation of gene expression. *Current Topics in Microbiology and Immunology*, 326:277-290.

- Rose, A. B., Elfersi, T., Parra, G. and Korf, I. (2008). Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. *Plant Cell*, 20:543-551.

- Rose, A. B., Emami, S., Bradnam, K. and Korf, I. (2011). Evidence for a DNA-based mechanism of intron-mediated enhancement. *Frontiers in Plant Science*, 2:98.

- Rose, R. J. and Nolan, K. E. (2006). Genetic regulation of somatic embryogenesis with particular reference to *Arabidopsis thaliana* and *Medicago truncatula*. *In Vitro Cellular & Developmental Biology-Plant*, 42:473-481.

- Rose, R. J., Mantiri, F. R., Kurdyukov, S., Chen, S.-K., Wang, X.-D., Nolan, K. E. and Sheahan, M. B. (2010). Developmental biology of somatic embryogenesis. In E. C. Pua and M. R. Davey (Eds.), *Plant developmental biology - Biotechnological perspectives*. (pp. 3-26). Berlin: Springer.

- Rounsley, S. D., Ditta, G. S. and Yanofsky, M. F. (1995). Diverse roles for MADS box genes in *Arabidopsis* development. *The Plant Cell*, 7(8):1259-1269.

- Sablowski, R. (2007). The dynamic plant stem cell niches. *Current Opinion in Plant Biology*, 10:639-644.

- Sáenz-Laín, C. (1981). Research on *Daucus* (*Umbelliferae*). *Anales del Jardín Botánico de Madrid*, 37:481-533.

- Saisho, D., Nakazono, M., Lee, K. H., Tsutsumi, N., Akita, S. and Hirai, A. (2001). The gene for alternative oxidase-2 (AOX2) from *Arabidopsis thaliana* consists of five exons unlike other AOX genes and is transcribed at an early stage during germination. *Genes & Genetic Systems*, 76(2):89-97.

- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406-425.

- Santos-Macedo, E., Cardoso, H. G., Hernández, A., Peixe, A. A., Polidoros, A., Ferreira, A. and Arnholdt-Schmitt, B. (2009). Physiologic responses and gene diversity indicate olive alternative oxidase as a potential source for markers involved in efficient adventitious root induction. *Physiologia Plantarum*, 137(4):532-552.

- Santos, A. L. W., Silveira, V., Steiner, N., Vidor, M. and Guerra, M. P. (2002). Somatic embryogenesis in parana pine (*Araucaria angustifolia* (Bert.) O. Kuntze). *Brazilian Archives of Biology and Technology*, 45(1):97-106.

- Satoh, S., Kamada, H., Harada, H. and Fujii, T. (1986). Auxin-controlled glycoprotein release into the medium of embryogenic carrot cells. *Plant Physiology*, 81(3):931-933.

- Schauer, S. E., Schlüter, P. M., Baskar, R., Gheyselinck, J., Bolaños, A., Curtis, M. D. and Grossniklaus, U. (2009). Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. *The Plant Journal*, 59(6):987-1000.

- Scheckhuber, C. Q., Houthoofd, K., Weil, A. C., Werner, A., De Vreese, A., Vanfleteren, J. R. and Osiewacz, H. D. (2011). Alternative oxidase dependent respiration leads to an increased mitochondrial content in two long-lived mutants of the ageing model *Podospora anserina*. *PLoS ONE*, 6(1):e16620.

- Schmidt, E. D. L., Guzzo, F., Toonen, M. A. J. and De Vries, S. C. (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development*, 124:2049-2062.

- Schueler-Furman, O. and Baker, D. (2003). Conserved residue clustering and protein structure prediction. *Proteins: Structure, Function, and Bioinformatics*, 52(2):225-235.

- Schultz, T. F., Spiker, S. and Quatrano, R. S. (1996). Histone H1 enhances the DNA binding activity of the transcription factor EmBP-1. *Journal of Biological Chemistry*, 271(42):25742-25745.

- Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T. and Satou, M. (2002a). Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Functional & Integrative Genomics*, 2(6):282-291.

- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. (2002b). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal*, 31:279-292.

- Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S. A., Silman, I., Sussman, J. L., Futerman, A. H. and Aviezer, D. (2007). Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnology Journal*, 5:579-590.

- Shahmuradov, I. A. and Solovyev, V. V. (2015). Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinformatics*, 31(21):3544-3545.

- Shahmuradov, I. A., Kolchanov, N. A., Solovyev, V. V. and Ratner, V. A. (1986). Enhancer-like structures in moderately repetitive sequences of eukaryotic genomes. *Genetika*, 22(3):357-367.

- Sharifi, G., Ebrahimzadeh, H., Ghareyazie, B., Gharechahi, J. and Vatankhah, E. (2012). Identification of differentially accumulated proteins associated with embryogenic and non-embryogenic *calli* in saffron (*Crocus sativus* L.). *Proteome Science*, 10:3-10.

- Sharma, V. K., Hänsch, R., Mendel, R. R. and Schulze, J. (2005). Mature embryo axis-based high frequency somatic embryogenesis and plant regeneration from multiple cultivars of barley (*Hordeum vulgare* L.). *Journal of Experimental Botany*, 56:1913-1922.

- Sharp, W. R., Sondahl, M. R., Caldas, L. S. and Maraffa, S. B. (1980). The physiology of *in vitro* asexual embryogenesis. *Horticultural Review*, 2:268-310.

- Shen, Q. J., Casaretto, J. A., Zhang, P. and Ho, T. H. D. (2004). Functional definition of ABA-response complexes: the promoter units necessary and sufficient for ABA induction of gene expression in barley (*Hordeum vulgare* L.). *Plant Molecular Biology*, 54(1):111-124.

- Shibukawa, T., Yazawa, K., Kikuchi, A. and Kamada, H. (2009). Possible involvement of DNA methylation on expression regulation of carrot LEC1 gene in its 5′-upstream region. *Gene*, 437(1):22-31.

- Shim, S. I. and Jørgensen, R. B. (2000). Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. *Theoretical and Applied Genetics*, 101:227-233.

- Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Current Opinion in Plant Biology*, 3(3):217-223.

- Siedow, J. N., Umbach, A. L. and Moore, A. L. (1995). The active site of the cyanide-resistant oxidase from plant mitochondria contains a binuclear iron center. *FEBS Letters*, 362(1):10-14.

- Sieger, S. M., Kristensen, B. K., Robson, C. A., Amirsadeghi, S., Eng, E. W., Abdel-Mesih, A. and Vanlerberghe, G. C. (2005). The role of alternative oxidase in modulating carbon use efficiency and growth during macronutrient stress in tobacco cells. *Journal of Experimental Botany*, 56(416):1499-1515.

- Simon, P. W. (1984). Carrot genetics. *Molecular Biology Reports*, 2:54-63.

- Simon, P. W. (2000). Domestication, historical development, and modern breeding of carrot. *Plant Breeding Reviews*, 19:157-190.

- Simons, B. H., Millenaar, F. F., Mulder, L., Van Loon, L. C. and Lambers, H. (1999). Enhanced expression and activation of the alternative oxidase during infection of *Arabidopsis* with *Pseudomonas syringae* pv tomato. *Plant Physiology*, 120:529-538.

- Smertenko, A. and Bozhkov, P. (2014). The life and death signalling underlying cell fate determination during somatic embryogenesis. In P. Nick and Z. Opatrny (Eds.), *Applied Plant Cell Biology*. (pp. 131-178). Heidelberg: Springer-Verlag.

- Sokal R. R. and Michener C. D. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Scientific Bulletin*, 38:1409-1438.

- Solovyev, V. (2002). Structure, properties and computer identification of eukaryotic genes. In T. Lengauer (Ed.), *Bioinformatics - From genomes to drugs.* (pp. 59-111). Weinheim: Wiley-VCH Verlag GmbH & Co.

- Solovyev, V. V and Kolchanov, N. A. (1994). Search for functional sites using consensus. In N. A. Kolchanov and H. A. Lim (Eds.), *Computer analysis of*

genetic macromolecules: structure, function and evolution. (pp. 16-21). Singapore: World Scientific.

- Soufflet-Freslon, V., Jourdan, M., Clotault, J., Huet, S., Briard, M., Peltier, D. and Geoffriau, E. (2013). Functional gene polymorphism to reveal species history: The case of the CRTISO gene in cultivated carrots. *PLoS ONE*, 8(8):e70801.

- Souter, M. and Lindsey, K. (2000). Polarity and signalling in plant embryogenesis. *Journal of Experimental Botany*, 51:971-983.

- Spalik, K. and Downie, S. R. (2007). Intercontinental disjunctions in *Cryptotaenia* (*Apiaceae*, *Oenantheae*): an appraisal using molecular data. *Journal of Biogeography*, 34:2039-2054.

- St. Pierre, M. D. and Bayer, R. J. (1991). The impact of domestication on the genetic variability in the orange carrot, cultivated *Daucus carota* ssp. *sativus* and the genetic homogeneity of various cultivars. *Theoretical and Applied Genetics*, 82:249-253.

- Stein, M. and Nothnagel, T. (1995). Some remarks on carrot breeding (*Daucus carota sativus* Hoffm.). *Plant Breeding*, 114:1-11.

- Stenmark, P. and Nordlund, P. (2003). A prokaryotic alternative oxidase present in the bacterium *Novosphingobium aromaticivorans*. *FEBS Letters*, 552:189-192.

- Steward, F. C., Mapes, M. O. and Smith, J. (1958). Growth and organized development of cultured cells. I. Growth and division of freely suspended cells. *American Journal of Botany*, 45:693-703.

- Stolarczyk, J. and Janick, J. (2011). Carrot: History and iconography. *Chronica Horticulturae*, 51:13-18.

- Sugie, A., Murai, K. and Takumi, S. (2007). Alteration of respiration capacity and transcript accumulation level of alternative oxidase genes in necrosis lines of common wheat. *Genes & Genetic Systems*, 82(3):231-239.

- Sugie, A., Naydenov, N., Mizuno, N., Nakamura, C. and Takumi, S. (2006). Overexpression of wheat alternative oxidase gene Waox1a alters respiration capacity and response to reactive oxygen species under low temperature in transgenic *Arabidopsis*. *Genes Genet. Systems*, 81:349-354.

- Sujatha, R. (2011). Pros and cons of somatic embryogenesis for mass multiplication of perennial plantation crops. In D. Hemming (Ed.), *Plant Sciences Reviews*. (pp. 1-20). Oxfordshire: CABI Publishing.

- Sun, Q., Sun, H., Bell, R. L., Li, H. and Xin, L. (2011). Variation of phenotype, ploidy level, and organogenic potential of *in vitro* regenerated polyploids of *Pyrus communis*. *Plant Cell, Tissue and Organ Culture*, 107:131-140.

- Suprasanna, P. and Bapat, V. A. (2005). Differential gene expression during somatic embryogenesis. In A. Mujib and J. Šamaj (Eds.), *Somatic Embryogenesis (Plant Cell Monographs)*. (pp. 305-320). Berlin: Springer.

- Swain, S. M., Reid, J. B. and Kamiya, Y. (1997). Gibberellins are required for embryo growth and seed development in pea. *Plant Journal*, 12:1329-1338.

- Sweetlove, L. J., Fait, A., Nunes-Nesi, A., Williams, T. and Fernie, A. R. (2007). The mitochondrion: an integration point of cellular metabolism and signalling. *Critical Reviews in Plant Sciences*, 26:17-43.

- Takahata, K. (2008). Isolation of carrot Argonaute1 from subtractive somatic embryogenesis cDNA library. *Bioscience, Biotechnology and Biochemistry*, 72(3):900-904.

- Takaichi, M. and Oeda, K. (2000). Transgenic carrots with enhanced resistance against two major pathogens, *Erysipheheraclei* and *Alternariadauci*. *Plant Science*, 153(2):135-144.

- Tamura, K., Nei, M. and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences*, 101:11030-11035.

- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10):2731-2739.

- Tangolar, S. G., Büyükalaca, S. and Ergenoğlu, F. (2008). High efficiency somatic embryogenesis from immature zygotic embryos of grapevine: the effect of genotype, media, 2,4-D, and incubation conditions. *Turkish Journal of Agriculture and Forestry*, 32(4):311-317.

- Tavares, A. C., Salgueiro, L. R. and Canhoto, J. M. (2010). *In vitro* propagation of the wild carrot *Daucus carota* L. subsp. *halophilus* (Brot.) A. Pujadas for conservation purposes. *In Vitro Cellular & Developmental Biology-Plant*, 46(1), 47-56.

- Thi, L. T. and Pleschka, E. (2005). Somatic embryogenesis of some *Daucus* species influenced by ABA. *Journal of Applied Botany and Food Quality*, 79(1):1-4.

- Thomas, C. and Jiménez, V. M. (2005). Mode of action of plant hormones and plant growth regulators during induction of somatic embryogenesis: molecular aspects. *Plant Cell Monographs*, 2:157-175.

- Thomas, J. C., Guiltinan, M. J., Bustos, S., Thomas, T. and Nessler, C. (1989). Carrot (*Daucus carota*) hypocotyl transformation using *Agrobacterium tumefaciens*. *Plant Cell Reports*, 8(6):354-357.

- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673-4680.

- Torné, J. M., Moysset, L., Santos, M. and Simon, E. (2001). Effects of light quality on somatic embryogenesis in *Araujia sercifera*. *Physiologia Plantarum*, 111:405-411.

- Umbach, A. L. and Siedow, J. N. (1993). Covalent and noncovalent dimers of the cyanide-resistant alternative oxidase protein in higher plant mitochondria and their relationship to enzyme activity. *Plant Physiology*, 103(3):845-854.

- Umbach, A. L., Fiorani, F. and Siedow, J. N. (2005). Characterization of transformed *Arabidopsis* with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. *Plant Physiology*, 139(4):1806-1820.

- Umehara, M., Eguchi, I., Kaneko, D., Ono, M. and Kamada, H. (2005). Evaluation of gene flow and its environmental effects in the field. *Plant Biotechnology*, 22:497-504.

- Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proceedings of the National Academy of Sciences*, 97(21):11632-11637.

- Vanisree, M., Chen. Y. L., Shu-Fung, L., Satish, M. N., Chien, Y. L. and Hsin-Sheng, T. (2004). Studies on the production of some important secondary metabolites from medicinal plants by plant tissue cultures. *Botanical Bulletin of Academia Sinica*, 45:1-22.

- Vanlerberghe, G. C. and McIntosh, L. (1996). Signals regulating the expression of the nuclear gene encoding alternative oxidase of plant mitochondria. *Plant Physiology*,111:589-595.

- Vanlerberghe, G. C. and McIntosh, L. (1997). Alternative oxidase: From gene to function. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48:703-734.

- Vanlerberghe, G. C., Cvetkovska, M. and Wang, J. (2009). Is the maintenance of homeostatic mitochondrial signaling during stress a physiological role for alternative oxidase?. *Physiologia Plantarum*, 137:392-406.

- Vanlerberghe, G. C., Robson, C. A. and Yip, J. Y. (2002). Induction of mitochondrial alternative oxidase in response to a cell signal pathway down-regulating the cytochrome pathway prevents programmed cell death. *Plant Physiology*, 129:1829-1842.

- Vanlerberghe, G. C., Vanlerberghe, A. E. and McIntosh, L. (1994). Molecular genetic alteration of plant respiration: silencing and overexpression of alternative oxidase in transgenic tobacco. *Plant Physiology*, 106:1503-1510.

- Vasil, I. K. (1987). Developing cell and tissue culture systems for the improvement of cereal and grass crops. *Journal of Plant Physiology*, 128:193-218.

- Veit, B. (2006). Stem cell signalling networks in plants. *Plant Molecular Biology*, 60:793-810.

- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L, Rusch, D., *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304:66-74.

- Verdeil, J.-L., Alemanno, L., Niemenak, N. and Tranbarger, T. J. (2007). Pluripotent versus totipotent plant stem cells: dependence versus autonomy?. *Trends Plant Science*, 12:245-252.

- Vivek, B. S. and Simon, P. W. (1999). Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). *Theoretical and Applied Genetics*, 99:58-64.

- Vries, S. C., Booij, H., Meyerink, P., Huisman, G., Wilde, D. H., Thomas, T. L. and Van Kammen, A. (1988). Acquisition of embryogenic potential in carrot cell-suspension cultures. *Planta*, 176:196-204.

- Wang, H., Caruso, L. V., Downie, A. B. and Perry, S. E. (2004). The embryo MADS domain protein AGAMOUS-like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. *Plant Cell*, 16:1206-1219.

- Wang, H., Huang, J., Liang, X. and Bi, Y. (2012). Involvement of hydrogen peroxide, calcium, and ethylene in the induction of the alternative pathway in chilling-stressed *Arabidopsis callus*. *Planta*, 235:53-67.

- Wang, J. and Vanlerberghe, G. C. (2013). A lack of mitochondrial alternative oxidase compromises capacity to recover from severe drought stress. *Physiologia Plantarum*, 149(4):461-473.

- Whelan, J., Millar, A. H. and Day, D. A. (1996). The alternative oxidase is encoded in a multigene family in soybean. *Planta*, 198:197-201.

- Wilde, H. D., Nelson, W. S., Booij, H., De Vries S. C. and Thomas, T. L. (1988). Gene-expression programs in embryogenic and non-embryogenic carrot cultures. *Planta*, 176:205-211.

- Wu, X., Ren, C., Joshi, T., Vuong, T., Xu, D. and Nguyen, H. T. (2010). SNP discovery by high-throughput sequencing in soybean. *BMC Genomics*, 11(1):1.

- Yamaki, S., Ohyanagi, H., Yamasaki, M., Eiguchi, M., Miyabayashi, T., Kubo, T., and Nonomura, K. I. (2013). Development of INDEL markers to discriminate

all genome types rapidly in the genus *Oryza*. *Breeding Science*, 63(3):246-254.

- Yang, X. and Zhang, X. (2010). Regulation of somatic embryogenesis in higher plants. *Critical Reviews in Plant Sciences*, 29:36-57.

- Yeung, E. C. (1995). Structural and developmental patterns in somatic embryogenesis. In T. A. Thorpe (Ed.), *In vitro embryogenesis in plants*. (pp. 203-247). London: Kluwer Academic Publishers.

- Yildiz, M. (2013). Plant responses at different ploidy levels. In M. Silva-Opps (Ed.), *Current Progress in Biological Research*. (pp. 363-385). Rijeka: InTech.

- Yip, J. Y. H. and Vanlerberghe, G. C. (2001). Mitochondrial alternative oxidase acts to dampen the generation of active oxygen species during a period of rapid respiration induced to support a high rate of nutrient uptake. *Physiologia Plantarum*, 112:327-333.

- Yu, M., Yuan, M. and Ren, H. (2006). Visualization of actin cytoskeletal dynamics during the cell cycle in tobacco (*Nicotiana tabacum* L. cv Bright Yellow) cells. *Biology of the Cell*, 98(5):295-306.

- Zavattieri, M. A., Frederico, A. M., Lima, M., Sabino, R. and Arnholdt-Schmitt, B. (2010). Induction of somatic embryogenesis as an example of stress-related plant reactions. *Electronic Journal of Biotechnology*, 13.

- Zeng, F., Zhang, X., Cheng, L., Hu, L., Zhu, L., Cao, J. and Guo, X. (2007). A draft gene regulatory network for cellular totipotency reprogramming during plant somatic embryogenesis. *Genomics*, 90(5):620-628.

- Zhang, J. E., Wen-Wu, G. U. O. and Xiu-Xin, D. E. N. G. (2006). Relationship between ploidy variation of *Citrus calli* and competence for somatic embryogenesis. *Acta Genetica Sinica*, 33(7):647-654.

- Zhang, W., Ruan, J., Ho, T. H. D., You, Y., Yu, T. and Quatrano, R. S. (2005). Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid-and abiotic stress-responsive genes in *Arabidopsis thaliana*. *Bioinformatics*, 21(14):3074-3081.

- Zhiponova, M. K., Pettkó-Szandtner, A., Stelkovics, É., Neer, Z., Bottka, S., Krenács, T. and Szilák, L. (2006). Mitosis-specific promoter of the alfalfa cyclin-dependent kinase gene (Medsa; CDKB2; 1) is activated by wounding and ethylene in a non-cell division-dependent manner. *Plant Physiology*, 140(2):693-703.

- Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology*, 53:247.

- Zimmerman, J. (1993). Somatic embryogenesis: A model for early development in higher plants. *The Plant Cell*, 5(10):1411-1423.

- Zuckerkandl, E. and Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2):357-366.

APPENDICES

Appendix 1 - Data compilation, concerning cell lines *calli* characterization, collected during induction, establishment and SE efficiency phenotyping sets. **Accession** - Number identifying the accession and used as the first cell line code number identifier (see Table 2.1 and section 2.2.1); **Embryo amount and quality** - Cell line qualitative classification according to the qualitative table (see Appendix 2, section 2.2.2 and 3.2); **Grey lines** - Cell lines identified as stable at the end of the first phenotyping set (see section 3.2.1) and used for genetic diversity evaluation using the cTBP method (see section 3.2.3). Cell lines used during the second phenotyping set (see section 3.2.3); **Number** - Individual *calli* number identifying it within the accession and used as the second cell line code number identifier (see section 2.2.1); **Source tissue** - Tissue from where *calli* developed.

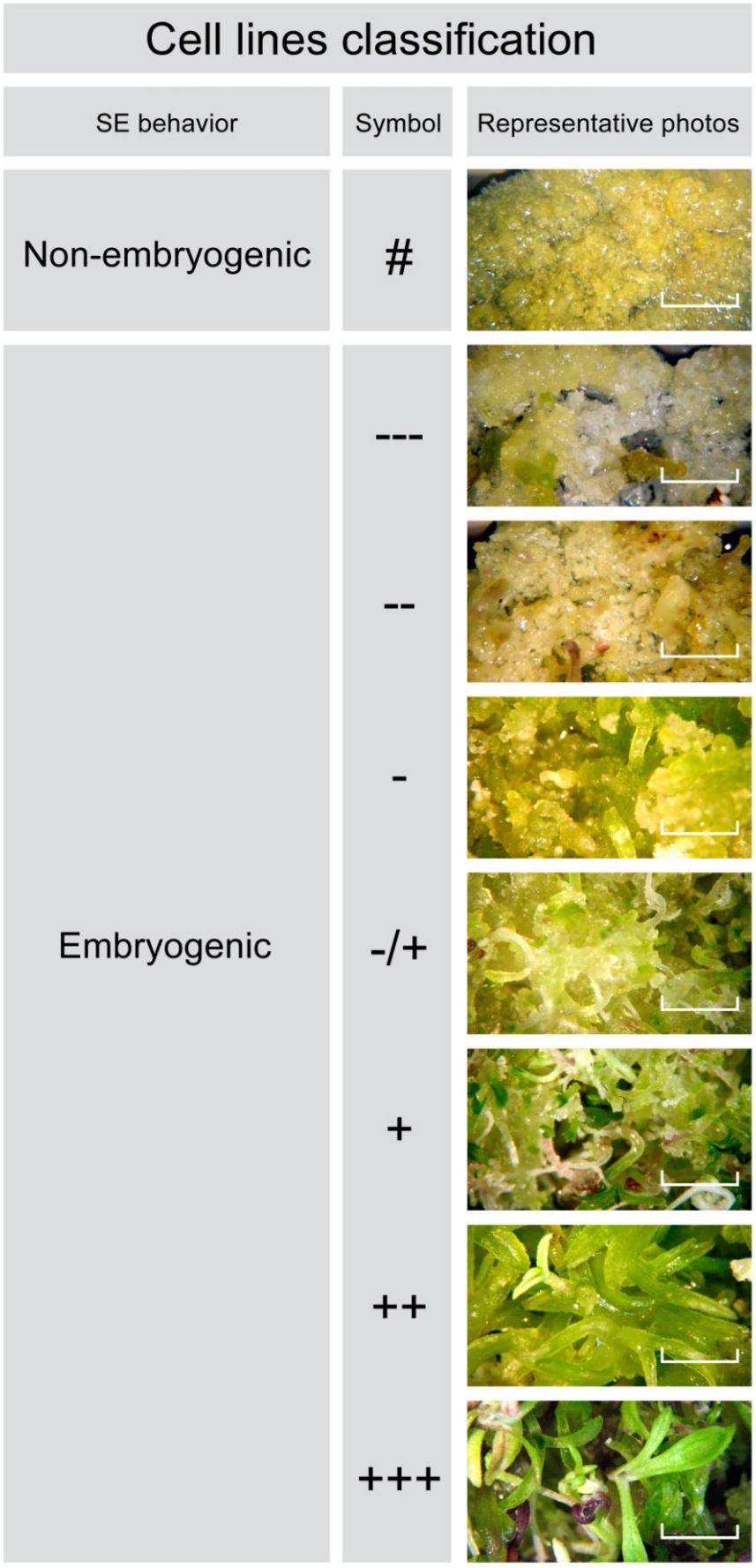
Appendix 1

Cell Line				First phenotyping set (4 replicates)								Second phenotyping set (4 replicates)							
				Relative frequency					Classification			Relative frequency					Classification		
Accession	Number	Source tissue	Characterization	Very efficient	Efficient	Inefficient	Very inefficient	Non-embryogenic	Stability	Efficiency	Embryo amount and quality	Very efficient	Efficient	Inefficient	Very inefficient	Non-embryogenic	Stability	Efficiency	Embryo amount and quality
1	1	Seedling	Friable, yellow and firm	0,50	0,25	0,25	0,00	0,00	Instable	Very efficient / efficient	+	Discarded					Discarded		
1	2	Seedling	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	++	Not used					Not used		
1	3	Seed internal tissues	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +	Discarded					Discarded		
1	4	Seedling	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient	+++
1	5	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,25	0,00	Instable	Efficient / inefficient	- / +								
1	6	Seedling	Friable, yellow / dark and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-	Discarded					Discarded		
1	7	Seedling	Friable, yellow and firm	0,75	0,00	0,25	0,00	0,00	Instable	Very efficient / efficient	-								
2	1	Seedling	Friable, yellow / white and firm	0,50	0,00	0,50	0,00	0,00	Instable	Very efficient / inefficient	- - -	Discarded					Discarded		
2	2	Seedling	Friable, yellow / white and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
2	3	Seedling	Friable, yellow / white and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	Not used					Not used		
2	4	Seed internal tissues	Friable, yellow and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#
2	5	Seed internal tissues	Friable, yellow / white and firm	0,25	0,00	0,25	0,50	0,00	Instable	Inefficient / very inefficient	- - -	Discarded					Discarded		
2	6	Seedling	Friable, yellow / white and firm	0,00	0,25	0,75	0,00	0,00	Stable	Efficient / inefficient	-	Not used					Not used		
2	7	Seedling	Friable, yellow / white and viscous	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#								
2	8	Seedling	Friable, yellow and firm	0,25	0,00	0,50	0,00	0,25	Instable	Inefficient / non-embryogenic	- -	Discarded					Discarded		
3	1	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,00	0,50	0,50	Instable	Very inefficient / non-embryogenic	- - -								
3	2	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,00	0,25	0,75	Instable	Very inefficient / non-embryogenic	- - -	Discarded					Discarded		
3	3	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,00	0,50	0,50	Instable	Very inefficient / non-embryogenic	- - -								
3	4	Seedling	Friable, yellow / dark and firm	0,00	0,50	0,00	0,25	0,25	Instable	Very inefficient / non-embryogenic	- - -								
3	5	Seed internal tissues	Friable, yellow / dark and firm	0,00	0,00	0,75	0,25	0,00	Stable	Inefficient	- - -	0,00	0,00	0,00	1,00	0,00	Stable	Very inefficient	- - -
3	6	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,50	0,50	0,00	Stable	Inefficient	+	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	+
3	7	Seedling	Friable, dark and viscous	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#
3	8	Seedling	Friable, dark and viscous	0,00	0,25	0,50	0,00	0,25	Instable	Inefficient / non-embryogenic	- -	Discarded					Discarded		
4	1	Seedling	Friable, yellow and firm	0,25	0,50	0,25	0,00	0,00	Stable	Efficient	- / +	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	-
4	2	Seedling	Friable, yellow and firm	0,00	0,00	0,50	0,00	0,50	Instable	Inefficient / non-embryogenic	- - -								
4	3	Seedling	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Instable	Efficient	- -	Discarded					Discarded		
4	4	Seedling	Friable, yellow / white and firm / viscous	0,00	0,25	0,25	0,50	0,00	Instable	inefficient / very inefficient	- - -								
4	5	Seedling	Friable, yellow and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#
4	6	Seed internal tissues	Friable, yellow and firm	0,00	0,25	0,00	0,25	0,50	Instable	Efficient / non-embryogenic	- - -	Discarded					Discarded		
4	7	Seedling	Friable, yellow / dark and firm	0,00	0,25	0,00	0,00	0,75	Instable	Efficient / non-embryogenic	- - -								
4	8	Seedling	Friable, yellow and firm	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	- / +	Not used					Not used		
5	1	Seedling	Friable, yellow and firm / viscous	0,25	0,50	0,00	0,00	0,25	Instable	Efficient / non-embryogenic	- -								
5	2	Seedling	Friable, yellow and firm / viscous	0,00	0,50	0,00	0,00	0,50	Instable	Efficient / non-embryogenic	- -	Discarded					Discarded		
5	3	Seedling	Friable, yellow and firm / viscous	0,00	0,75	0,00	0,25	0,00	Instable	Efficient / very inefficient	- - -								
5	4	Seed internal tissues	Friable, yellow and firm / viscous	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	+	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	+
5	5	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,00	0,25	Instable	Efficient / non-embryogenic	- / +								
5	6	Seedling	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Instable	Efficient / inefficient	- -	Discarded					Discarded		
5	7	Seedling	Friable, yellow and firm	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	- / +	Not used					Not used		
5	8	Seedling	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Instable	Efficient / inefficient	- -	Discarded					Discarded		
6	1	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,00	0,25	Instable	Efficient / non-embryogenic	- - -								
6	2	Seedling	Friable, yellow / white and firm	0,50	0,00	0,50	0,00	0,00	Instable	Very efficient / inefficient	- -	Discarded					Discarded		
6	3	Seedling	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-
6	4	Seedling	Friable, yellow and firm	0,25	0,50	0,25	0,00	0,00	Instable	Efficient / inefficient	- -								
6	5	Seedling	Friable, yellow and firm / viscous	0,25	0,25	0,25	0,25	0,00	Instable	Very efficient / inefficient	- - -	Discarded					Discarded		
6	6	Seedling	Friable, white and firm	0,00	0,00	0,25	0,75	0,00	Stable	Very inefficient	- - -	0,00	0,00	0,00	1,00	0,00	Stable	Very inefficient	- - -
7	1	Seedling	Friable, yellow and firm / viscous	0,00	0,00	0,00	0,25	0,75	Instable	Inefficient / non-embryogenic	- - -	Discarded					Discarded		
7	2	Seedling	Friable, dark and viscous	0,25	0,50	0,00	0,00	0,25	Instable	Eficient / non-embryogenic	- - -								
7	3	Seedling	Friable, yellow and firm / viscous	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,25	0,75	Instable	Very inefficient / non-embryogenic	#
7	4	Seedling	Indured, yellow and firm	0,00	0,00	0,00	0,50	0,50	Instable	Very inefficient / non-embryogenic	- - -								
7	5	Seed internal tissues	Friable, yellow and firm	0,00	0,00	0,25	0,00	0,75	Instable	Inefficient / non-embryogenic	- - -	Discarded					Discarded		
7	6	Seedling	Friable, yellow and firm	0,00	0,00	0,00	0,25	0,75	Instable	Very inefficient / non-embryogenic	- - -								
7	7	Seedling	Friable, yellow and viscous	0,00	0,00	0,00	0,25	0,75	Instable	Very inefficient / non-embryogenic	- - -								
7	8	Seedling	Friable, yellow and firm	0,00	0,00	0,00	0,75	0,25	Stable	Very inefficient	- -	Not used					Not used		
8	1	Seed internal tissues	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	+	Discarded					Discarded		
8	2	Seedling	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++
8	3	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	+								
8	4	Seedling	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	++	Discarded					Discarded		
8	5	Seedling	Friable, yellow / green and firm	0,00	0,75	0,25	0,00	0,00	Instable	Efficient / inefficient	-								
9	1	Seedling	Friable, white and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +								
9	2	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-								
9	3	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	-								
9	4	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +	Discarded					Discarded		
9	5	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +								
9	6	Seedling	Friable, yellow and firm	0,25	0,25	0,50	0,00	0,00	Instable	Very efficient / inefficient	- -								
9	7	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
10	1	Seedling	Friable, yellow and firm	0,50	0,25	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	- - -								
10	2	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
10	3	Seedling	Friable, yellow and firm	0,00	0,50	0,00	0,00	0,50	Instable	Efficient / non-embryogenic	- - -	Discarded					Discarded		
10	4	Seedling	Friable, yellow and firm	0,00	0,50	0,50	0,00	0,00	Instable	Efficient / inefficient	- - -								
10	5	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	- -								
10	6	Seedling	Friable, yellow and firm	0,25	0,75	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-
10	7	Seedling	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Instable	Efficient / inefficient	- - -	Discarded					Discarded		
11	1	Seedling	Friable, yellow / white and firm	0,00	0,25	0,00	0,00	0,75	Instable	Efficient / non-embryogenic	- - -	Discarded					Discarded		
11	2	Seedling	Friable, yellow / white and firm	0,00	0,50	0,50	0,00	0,00	Stable	Efficient / inefficient	- -	Not used					Not used		
11	3	Seedling	Friable, yellow / white and firm	0,00	0,00	0,25	0,00	0,75	Instable	Inefficient / non-embryogenic	- - -	Discarded					Discarded		
11	4	Seedling	Friable, yellow / white and firm	0,00	0,00	0,25	0,50	0,25	Stable	Inefficient / very inefficient	- - -	Not used					Not used		
12	1	Seed internal tissues	Friable, yellow / white and firm	0,00	0,00	0,00	0,50	0,50	Instable	Very inefficient / non-embryogenic	- - -								
12	2	Seedling	Friable, yellow and firm	0,50	0,25	0,25	0,00	0,00	Instable	Very efficient / inefficient	-	Discarded					Discarded		
12	3	Seedling	Friable, yellow / white and firm	0,25	0,75	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
12	4	Seedling	Friable, yellow and firm	0,00	0,25	0,00	0,00	0,75	Instable	Efficient / non-embryogenic	- - -								
13	1	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-	Discarded					Discarded		
13	2	Seedling	Friable, yellow and firm	0,25	0,25	0,25	0,25	0,00	Instable	Very efficient / very inefficient	- -								
13	3	Seedling	Friable, yellow and firm	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	- / +	0,25	0,75	0,00	0,00	0,00	Instable	Efficient	- / +
13	4	Seedling	Friable, yellow and firm	0,25	0,75	0,00	0,00	0,00	Stable	Efficient	-	Not used					Not used		
14	1	Seed internal tissues	Friable, yellow and firm	0,00	0,50	0,50	0,00	0,00	Stable	Efficient / inefficient	-	Not used					Not used		
14	2	Seed internal tissues	Friable, yellow / white and firm	0,50	0,25	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	- - -								
14	3	Seed internal tissues	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	- - -	Discarded					Discarded		
14	4	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- - -								
15	1	Seedling	Friable, yellow / white and firm	0,25	0,50	0,25	0,00	0,00	Instable	Very efficient / inefficient	- - -	Discarded					Discarded		
15	2	Seedling	Friable, yellow and firm / viscous	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	Not used					Not used		
15	3	Seedling	Friable, yellow and firm	0,25	0,50	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	- - -	Discarded					Discarded		
15	4	Seedling	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Stable	Efficient / inefficient	-	Not used					Not used		
16	1	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,00	0,25	0,75	Instable	Inefficient / non-embryogenic	- - -	Discarded					Discarded		
16	2	Seedling	Friable, yellow / green and viscous	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	Not used					Not used		
16	3	Seedling	Friable, yellow / white and firm	0,25	0,25	0,00	0,25	0,25	Instable	Very efficient / non-embryogenic	- - -	Discarded					Discarded		
17	1	Seedling	Friable, yellow and firm	0,50	0,25	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	-								
17	2	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- - -	Discarded					Discarded		
17	3	Seedling	Friable, yellow and firm	0,00	0,50	0,25	0,25	0,00	Instable	Efficient / inefficient	- - -								
17	4	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	+								
17	5	Seedling	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++	0,00	0,00	0,25	0,50	0,25	Instable	Very inefficient / non-embryogenic	- - -
17	6	Seedling	Friable, yellow / green and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	+	Discarded					Discarded		
18	1	Seedling	Friable, yellow / dark and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#
18	2	Seedling	Friable, white and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	Not used					Not used		
18	3	Seedling	Friable, dark / white and firm	0,00	0,75	0,25	0,00	0,00	Stable	Efficient / inefficient	- / +	0,00	0,75	0,25	0,00	0,00	Stable	Efficient / inefficient	-
18	4	Seedling	Friable, yellow and firm	0,25	0,25	0,25	0,00	0,25	Instable	Very efficient / non-embryogenic	- - -	Discarded					Discarded		
18	5	Seedling	Friable, yellow / white and firm	0,00	0,00	0,50	0,50	0,00	Stable	Inefficient / very inefficient	- - -	Not used					Not used		
18	6	Seedling	Friable, yellow / white / dark and firm	0,00	0,50	0,00	0,00	0,50	Instable	Efficient / non-embryogenic	- - -	Discarded					Discarded		
19	1	Seedling	Friable, white and firm	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#	0,00	0,00	0,00	0,00	1,00	Stable	Non-embryogenic	#
19	2	Seedling	Friable, white and firm	0,00	0,25	0,25	0,00	0,50	Instable	Efficient / non-embryogenic	- - -								
19	3	Seedling	Friable, white and firm	0,25	0,00	0,25	0,00	0,50	Instable	Very efficient / non-embryogenic	- - -	Discarded					Discarded		
19	4	Seedling	Friable, white and firm	0,25	0,25	0,25	0,00	0,25	Instable	Very efficient / non-embryogenic	- - -								
20	1	Seedling	Friable, white and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	+	Discarded					Discarded		
20	2	Seedling	Indured, white and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	+++
20	3	Seedling	Indured, white and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	++	Discarded					Discarded		
21	1	Seedling	Friable, white and firm / viscous	0,00	0,75	0,00	0,00	0,25	Instable	Efficient / non-embryogenic	-	Discarded					Discarded		
21	2	Seedling	Friable, white and firm / viscous	0,00	0,75	0,25	0,00	0,00	Instable	Efficient / inefficient	-								
23	1	Seedling	Friable, yellow / white and firm	0,00	0,00	0,25	0,75	0,00	Stable	Very inefficient	- -	Not used					Not used		
23	2	Seed internal tissues	Friable, yellow / white and firm	0,00	0,75	0,00	0,00	0,25	Instable	Efficient / non-embryogenic	- - -								
23	3	Seed internal tissues	Friable, white / dark and viscous	0,25	0,75	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
23	4	Seedling	Friable, white and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +	Discarded					Discarded		
23	5	Seedling	Friable, yellow / white and viscous	0,25	0,50	0,00	0,25	0,00	Instable	Very efficient / very inefficient	-								
23	6	Seedling	Friable, yellow and firm	0,00	0,75	0,00	0,25	0,00	Instable	Efficient / inefficient	- -								
26	1	Seed internal tissues	Friable, yellow and firm	0,00	0,75	0,25	0,00	0,00	Stable	Efficient / inefficient	- / +	Not used					Not used		
26	2	Seed internal tissues	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +								
26	3	Seed internal tissues	Friable, yellow and firm	0,25	0,25	0,25	0,00	0,25	Instable	Very efficient / non-embryogenic	- / +	Discarded					Discarded		
26	4	Seed internal tissues	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	-								
26	5	Seed internal tissues	Friable, yellow / dark and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	- -								
26	L5.S.R.	Seed internal tissues	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	- / +	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +
27	1	Seedling	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +	Discarded					Discarded		
27	2	Seedling	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +
27	3	Seedling	Friable, yellow and firm	0,50	0,50	0,00	0,00	0,00	Instable	Very efficient / efficient	- / +	Discarded					Discarded		
27	4	Seedling	Friable, yellow and firm	1,00	0,00	0,00	0,00	0,00	Stable	Very efficient	- / +	Not used					Not used		
27	5	Seedling	Friable, yellow / white and firm	0,25	0,75	0,00	0,00	0,00	Instable	Very efficient / efficient	-	Discarded					Discarded		
28	1	Seed internal tissues	Friable, yellow / green and firm	0,75	0,25	0,00	0,00	0,00	Instable	Very efficient / efficient	++	Discarded					Discarded		
28	2	Seed internal tissues	Friable, yellow and firm	0,75	0,25	0,00	0,00	0,00	Stable	Very efficient / efficient	- / +	0,25	0,50	0,00	0,00	0,25	Instable	Very efficient / non-embryogenic	-
28	3	Seed internal tissues	Friable, yellow / white and firm	0,00	0,50	0,25	0,25	0,00	Instable	Efficient / inefficient	-								
28	4	Seed internal tissues	Friable, yellow / white and firm	0,25	0,50	0,25	0,00	0,00	Instable	Efficient / inefficient	-	Discarded					Discarded		
28	5	Seed internal tissues	Friable, yellow / white and firm	0,00	0,50	0,25	0,25	0,00	Instable	Efficient / very inefficient	- -								
28	6	Seed internal tissues	Friable, white and firm	0,25	0,75	0,00	0,00	0,00	Instable	Very efficient / efficient	-								

Appendix 1

.

Appendix 2 - Qualitative scale created to refine cell lines SE efficiency phenotype selection at the phenotyping sets end (see Appendix 1). The scale was based on the observed amount number and quality of embryos present in the culture at the T60 documentation point (see section 2.2.2 and 3.2). **Bar** - 4 mm.



Appendix 3 - **A** - *AOX1*, **B** - *AOX2a* and **C** - *AOX2b.* Structural characterization of the *AOX* sequences amplified from the selected cell lines with extreme SE efficiency phenotypes (see section 2.2.6 and 3.3.1). **aa** - Amino acid; **bp** - Base pair; **Cell line** - 2_4 - *D. c*. 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed; **Grey line** - Sequence missing the typical *AOX2a* start, not used in the bioinformatic analyses; **Nr.** - Number of identical sequences; **ORF** - Open reading frame; **Protein** - Full length of the deduced amino acid sequence; **S>E** - Position of start and end of the region analyzed; **Total** - Full length of the amplified sequence; **UTR** - Untranslated region.

A

				Expressed regions (bp)						Intragenic regions (bp)				Partial 5'-UTR (bp)	Partial 3'-UTR (bp)	Total (bp)	ORF (bp)	Protein (aa)
				Exon 1		Exon 2		Exon 3		Intron 1		Intron 2						
		Nr.	Cell line	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length					
Clone	**50**	14	2_4	12-443	432	1074-1562	489	1736-1795	60	444-1073	630	1563-1735	173	11	19	1814	981	326
	51	3		12-413	402	1126-1614	489	1787-1846	60	414-1125	712	1615-1786	172	11	19	1865	951	316
	1	6	4_5	12-413	402	1126-1614	489	1787-1846	60	414-1125	712	1615-1786	172	11	19	1865	951	316
	4	7	8_2	12-404	393	1112-1600	489	1773-1832	60	405-1111	707	1601-1772	172	11	19	1851	942	313
	1	8	18_1	12-413	402	1050-1538	489	1711-1770	60	414-1049	636	1539-1710	172	11	19	1789	951	316
	4	7	19_1	12-413	402	1121-1609	489	1783-1842	69	414-1120	707	1610-1782	173	11	19	1861	951	316
	1	8	20_2	12-413	402	1121-1609	489	1783-1842	69	414-1120	707	1610-1782	173	11	19	1861	951	316
	1	7	26_L5.S.R	12-413	402	1126-1614	489	1787-1846	60	414-1125	712	1615-1786	172	11	19	1865	951	316
	1	12	27_2	12-404	393	1112-1600	489	1773-1832	60	405-1111	707	1601-1772	172	11	19	1851	942	313
	47	2		12-404	393	1105-1593	489	1766-1825	60	405-1104	700	1594-1765	172	11	19	1844	942	313
	52	3		12-404	393	1112-1600	489	1773-1832	60	405-1111	707	1601-1772	172	11	19	1851	942	313
	Total	77																

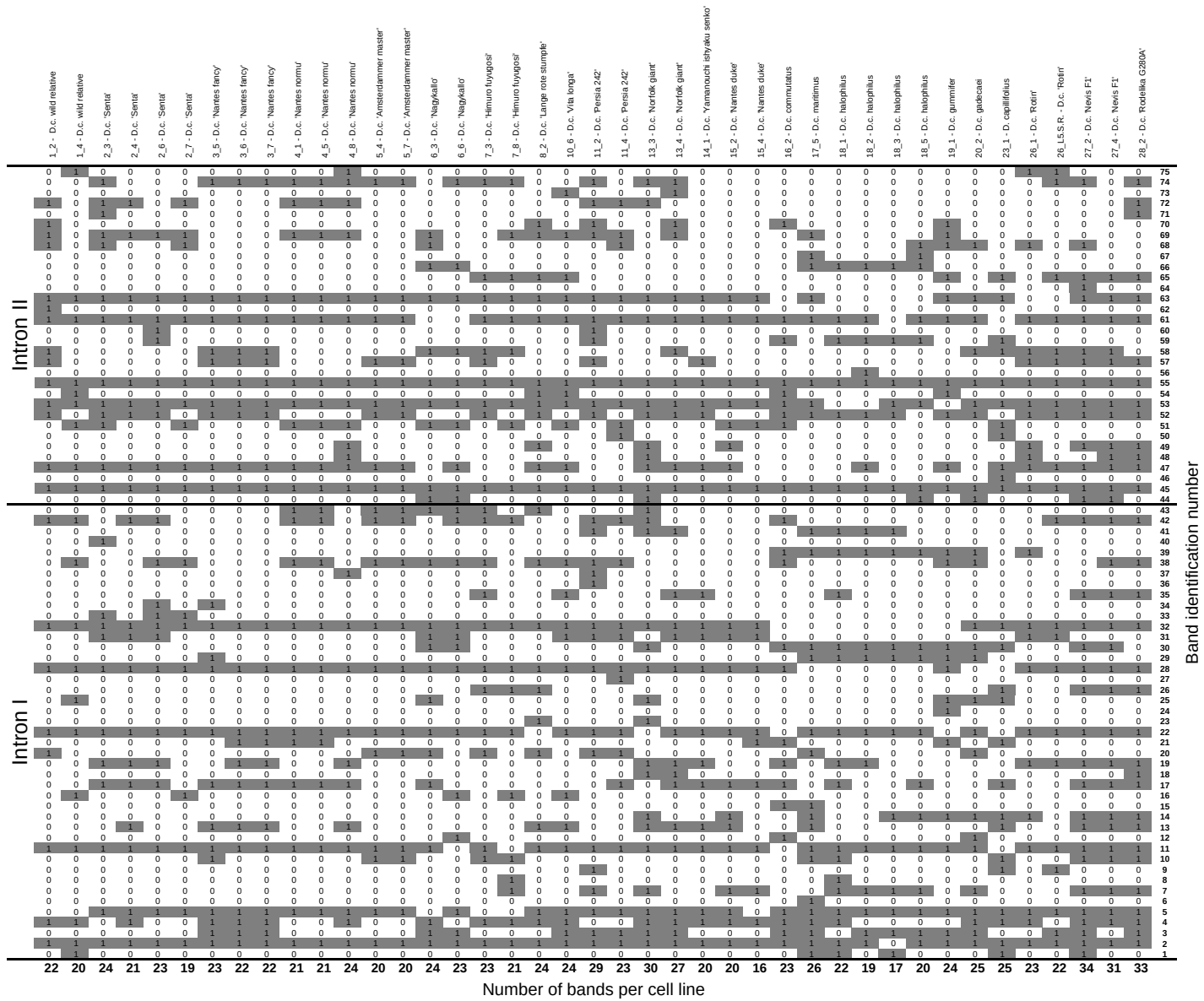
B

Clone		Nr.	Cell line	Expressed regions (bp)								Intragenic regions (bp)						Partial 3'-UTR (bp)	Total (bp)	ORF (bp)	Protein (aa)
				Exon 1		Exon 2		Exon3		Exon 4		Intron 1		Intron 2		Intron 3					
				S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length				
	47	8	**2_4**	1-339	339	2411-2539	129	3494-3982	489	4924-4983	60	340-2410	2071	2540-3493	954	3983-4923	941	36	5019	1017	338
	56	1		1-338	338	2410-2538	129	3493-3981	489	4923-4982	60	339-2409	2071	2539-3492	954	3982-4922	941	36	5018	1016	337
	8	9	**4_5**	1-339	339	2410-2538	129	3493-3981	489	4923-4982	60	340-2409	2070	2539-3492	954	3982-4924	941	36	5018	1017	338
	3	8	**8_2**	1-339	339	2410-2538	129	3493-3981	489	4923-4982	60	340-2409	2070	2539-3492	954	3982-4922	941	36	5018	1017	338
	17	1		1-339	339	2408-2536	129	3991-3979	489	4921-4980	60	340-2407	2068	2537-3490	954	3980-4920	941	36	5016	1017	338
	S1	10	**18_1**	1-339	339	2416-2544	129	3501-3989	489	4930-4989	60	340-2415	2076	2546-3500	956	3990-4929	940	36	5025	1017	338
	L56	7		1-339	339	2321-2449	129	3406-3894	489	5121-5180	60	340-2320	1981	2450-3405	956	3895-5120	1226	36	5216	1017	338
	L146	1		1-339	339	2297-2425	129	3382-3870	489	5097-5156	60	340-2296	1957	2426-3381	956	3871-5096	1226	36	5192	1017	338
	L177	2		1-339	339	2319-2447	129	3404-3892	489	5119-5178	60	340-2318	1979	2448-3405	956	3893-5118	1226	36	5214	1017	338
	L214	1		1-339	339	2320-2448	129	3405-3893	489	5120-5179	60	340-2319	1980	2449-3404	956	3894-5119	1226	36	5215	1017	338
	L219	3		1-339	339	2297-2425	129	3382-3870	489	5097-5156	60	340-2296	1957	2426-3381	956	3871-5096	1226	36	5192	1017	338
	4	4	**19_1**	1-336	336	2324-2452	129	3405-3893	489	4819-4878	60	337-2323	1987	2453-3404	952	3894-4818	925	36	4914	1014	337
	14	5		1-336	336	2326-2454	129	3407-3895	489	4821-4880	60	337-2325	1989	2455-3406	952	3896-4820	925	36	4916	1014	337
	64	3		1-336	336	2322-2450	129	3403-3891	489	4817-4876	60	337-2321	1985	2451-3402	952	3892-4816	925	36	4912	1014	337
	3	6	**20_2**	1-336	336	2324-2452	129	3405-3893	489	4819-4878	60	337-2323	1987	2453-3404	952	3894-4878	925	36	4914	1014	337
	7	1		1-335	335	2321-2449	129	3402-3890	489	4816-4875	60	336-2320	1985	2450-3401	952	3891-4815	925	36	4911	1013	336
	31	1		1-336	336	2321-2449	129	3402-3890	489	4818-4877	60	337-2320	1984	2450-3401	952	3891-4817	927	36	4913	1014	337
	35	2		1-336	336	2322-2450	129	3403-3891	489	4819-4878	60	337-2321	1985	2451-3402	952	3892-4818	927	36	4914	1014	337
	S5	11	**26_L5.S.R**	1-339	339	2410-2538	129	3493-3981	489	4923-4882	60	340-2409	2070	2539-3492	954	3982-4922	941	36	5018	1017	338
	L13	1		1-339	339	2410-2538	129	3496-3984	489	5211-5270	60	340-2409	2070	2539-3495	957	3985-5210	1226	36	5306	1017	338
	L101	3		1-339	339	2410-2538	129	3493-3981	489	5208-5267	60	340-2409	2070	2539-3492	954	3982-5207	1226	36	5303	1017	338
	L128	3		1-339	339	2401-2529	129	3487-3975	489	5202-5261	60	340-2400	2061	2530-3486	957	3976-5201	1226	36	5297	1017	338
	L142	4		1-339	339	2419-2547	129	3505-3993	489	5220-5279	60	340-2418	2079	2548-3504	957	3994-5219	1226	36	5315	1017	338
	27	8	**27_2**	1-339	339	2410-2538	129	3493-3981	489	4923-4982	60	340-2409	2070	2539-3492	954	3982-4922	941	36	5018	1017	338
	Total	103																			

C

				Expressed regions (bp)								Intragenic regions (bp)						Partial 5'-UTR (bp)	Partial 3'-UTR (bp)	Total (bp)	ORF (bp)	Protein (aa)
				Exon 1		Exon 2		Exon 3		Exon 4		Intron 1		Intron 2		Intron 3						
		Nr.	Cell line	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length	S>E	Length					
Clone	**57**	10	2_4	88-369	282	1389-1517	129	1609-2097	489	2183-2242	60	370-1388	1019	1518-1608	91	2098-2182	85	87	102	2344	960	319
	1_16	17	4_5	88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	3_21	13		88-369	282	1389-1517	129	1609-2097	489	2183-2242	60	370-1388	1019	1518-1608	91	2098-2182	85	87	102	2344	960	319
	0_32	3	8_2	88-369	282	929-1058	129	1150-1638	489	1724-1783	60	370-928	560	1059-1149	91	1639-1723	85	87	102	1885	960	319
	1_34	6		88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	1_45	6		88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	17	11	18_1	88-369	282	929-1058	129	1150-1638	489	1724-1783	60	370-928	560	1059-1149	91	1639-1723	85	87	102	1885	960	319
	2_26	3	19_1	88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	2_31	8		88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	5	11	20_2	85-366	282	1338-1517	129	1609-2097	489	2183-2242	60	367-1337	832	1518-1608	91	2098-2182	102	84	102	2171	960	319
	17	11	26_L5.S.R	88-369	282	1192-1320	129	1412-1900	489	1986-2045	60	367-1191	822	1321-1411	91	1901-1985	85	87	102	2147	960	319
	47	1	27_2	88-369	282	929-1058	129	1150-1638	489	1724-1783	60	370-928	560	1059-1149	91	1639-1723	85	87	102	1885	960	319
	56	7		88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	57	3		88-369	282	1193-1321	129	1413-1901	489	1987-2046	60	370-1192	823	1322-1412	91	1902-1986	85	87	102	2148	960	319
	Total	110																				

Appendix 4 - Presence (1) / absence (0) of cTBP markers (bands) detected on each one of the 41 selected cell lines selected from the first phenotyping set. The markers were used to construct the similarity matrix presented in the Appendix 5 (see sections 2.2.3 and 3.2.2). The cell line identification is presented in the Appendix 1.


1_2 - D.c. wild relative
1_4 - D.c. wild relative
2_3 - D.c. 'Senta'
2_4 - D.c. 'Senta'
2_6 - D.c. 'Senta'
2_7 - D.c. 'Senta'
3_5 - D.c. 'Nantes fancy'
3_6 - D.c. 'Nantes fancy'
3_7 - D.c. 'Nantes fancy'
4_1 - D.c. 'Nantes normu'
4_5 - D.c. 'Nantes normu'
4_8 - D.c. 'Nantes normu'
5_4 - D.c. 'Amsterdammer master'
5_7 - D.c. 'Amsterdammer master'
6_3 - D.c. 'Nagykallo'
6_6 - D.c. 'Nagykallo'
7_3 - D.c. 'Himuro fuyugosi'
7_8 - D.c. 'Himuro fuyugosi'
8_2 - D.c. 'Lange rote stumpfe'
10_6 - D.c. 'Vita longa'
11_2 - D.c. 'Persia 242'
11_4 - D.c. 'Persia 242'
13_3 - D.c. 'Norfolk giant'
13_4 - D.c. 'Norfolk giant'
14_1 - D.c. 'Yamanouchi ishyaku senko'
15_2 - D.c. 'Nantes duke'
15_4 - D.c. 'Nantes duke'
16_2 - D.c. commutatus
17_5 - D.c. maritimus
18_1 - D.c. halophilus
18_2 - D.c. halophilus
18_3 - D.c. halophilus
18_5 - D.c. halophilus
19_1 - D.c. gummifer
20_2 - D.c. gadecaei
23_1 - D. capillifolius
26_1 - D.c. 'Rotin'
26_1.5.S.R. - D.c. 'Rotin'
27_2 - D.c. 'Nevis F1'
27_4 - D.c. 'Nevis F1'
28_2 - D.c. 'Rodelika G280A'
Intron II
Intron I
Band identification number
22 20 24 21 23 19 23 22 22 21 21 24 20 20 24 23 23 21 24 24 29 23 30 27 20 20 16 23 26 22 19 17 20 24 25 25 23 22 34 31 33
Number of bands per cell line

Appendix 4

Appendix 5 - Similarity matrix constructed using the cTBP markers (see sections 2.2.3 and 3.2.2). The cell line identification is presented in the Appendix 1.

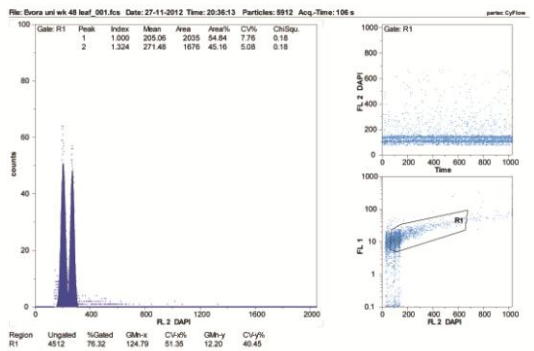
Appendix 5

	1_2 - D.c. wild relative	1_4 - D.c. wild relative	2_3 - D.c. 'Senta'	2_4 - D.c. 'Senta'	2_6 - D.c. 'Senta'	2_7 - D.c. 'Senta'	3_5 - D.c. 'Nantes fancy'	3_6 - D.c. 'Nantes fancy'	3_7 - D.c. 'Nantes fancy'	4_1 - D.c. 'Nantes normu'	4_5 - D.c. 'Nantes normu'	4_8 - D.c. 'Nantes normu'	5_4 - D.c. 'Amsterdammer master'	5_7 - D.c. 'Amsterdammer master'	6_3 - D.c. 'Nagykalo'	6_6 - D.c. 'Nagykalo'	7_3 - D.c. 'Himuro fuyugosi'	7_8 - D.c. 'Himuro fuyugosi'	8_2 - D.c. 'Lange rote stumpfe'	10_6 - D.c. 'Vita longa'	11_2 - D.c. 'Persia 242'	11_4 - D.c. 'Persia 242'	13_3 - D.c. 'Norfolk giant'	13_4 - D.c. 'Norfolk giant'	14_1 - D.c. 'Yamanouchi ishyaku senko'	15_2 - D.c. 'Nantes duke'	15_4 - D.c. 'Nantes duke'	16_2 - D.c. commutatus	17_5 - D.c. maritimus	18_1 - D.c. halophilus	18_2 - D.c. halophilus	18_3 - D.c. halophilus	18_5 - D.c. halophilus	19_1 - D.c. gummifer	20_2 - D.c. gadecaei	23_1 - D. capillifolius	26_1 - D.c. 'Rotin'	26_L5.S.R. - D.c. 'Rotin'	27_2 - D.c. 'Nevis F1'	27_4 - D.c. 'Nevis F1'	28_2 - D.c. 'Rodelika G280A'
1_2		0,619	0,652	0,744	0,622	0,683	0,667	0,739	0,682	0,609	0,667	0,609	0,708	0,698	0,609	0,489	0,711	0,605	0,652	0,565	0,667	0,667	0,538	0,653	0,667	0,571	0,579	0,444	0,500	0,364	0,390	0,308	0,381	0,522	0,596	0,383	0,667	0,636	0,607	0,604	0,582
1_4	0,619		0,545	0,634	0,605	0,718	0,558	0,545	0,571	0,682	0,651	0,636	0,565	0,634	0,591	0,605	0,605	0,634	0,591	0,727	0,490	0,605	0,520	0,511	0,600	0,650	0,667	0,512	0,435	0,381	0,359	0,324	0,350	0,500	0,533	0,489	0,558	0,571	0,519	0,549	0,528
2_3	0,652	0,545		0,800	0,766	0,791	0,638	0,708	0,696	0,708	0,766	0,750	0,640	0,622	0,583	0,553	0,511	0,533	0,542	0,625	0,604	0,723	0,556	0,706	0,727	0,682	0,650	0,468	0,440	0,435	0,465	0,341	0,455	0,500	0,490	0,408	0,638	0,652	0,586	0,545	0,632
2_4	0,744	0,634	0,800		0,818	0,700	0,727	0,800	0,791	0,711	0,727	0,800	0,723	0,667	0,578	0,545	0,591	0,571	0,667	0,711	0,640	0,727	0,667	0,792	0,878	0,780	0,703	0,545	0,553	0,512	0,500	0,368	0,439	0,489	0,522	0,478	0,682	0,698	0,655	0,692	0,704
2_6	0,622	0,605	0,766	0,818		0,714	0,652	0,681	0,667	0,681	0,696	0,638	0,653	0,727	0,553	0,565	0,565	0,500	0,596	0,638	0,692	0,696	0,528	0,680	0,744	0,651	0,615	0,565	0,449	0,489	0,524	0,400	0,465	0,511	0,500	0,417	0,609	0,667	0,561	0,630	0,607
2_7	0,683	0,718	0,791	0,700	0,714		0,571	0,605	0,585	0,791	0,810	0,698	0,667	0,650	0,605	0,619	0,524	0,600	0,605	0,744	0,583	0,762	0,490	0,565	0,615	0,667	0,629	0,429	0,444	0,341	0,421	0,333	0,462	0,558	0,545	0,409	0,571	0,537	0,491	0,520	0,538
3_5	0,667	0,558	0,638	0,727	0,652	0,571		0,851	0,889	0,596	0,609	0,681	0,694	0,773	0,553	0,565	0,696	0,591	0,596	0,638	0,577	0,565	0,604	0,760	0,791	0,698	0,615	0,478	0,612	0,533	0,524	0,450	0,512	0,511	0,625	0,542	0,652	0,711	0,737	0,704	0,714
3_6	0,739	0,545	0,708	0,800	0,681	0,605	0,851		1,000	0,667	0,681	0,750	0,680	0,667	0,583	0,553	0,638	0,578	0,667	0,667	0,642	0,596	0,630	0,863	0,818	0,682	0,650	0,596	0,560	0,478	0,512	0,390	0,455	0,583	0,571	0,531	0,681	0,739	0,724	0,691	0,702
3_7	0,682	0,571	0,696	0,791	0,667	0,585	0,889	1,000		0,652	0,667	0,739	0,667	0,698	0,565	0,578	0,667	0,558	0,609	0,652	0,588	0,578	0,654	0,816	0,857	0,714	0,684	0,578	0,542	0,500	0,537	0,410	0,476	0,522	0,596	0,553	0,711	0,773	0,750	0,717	0,727
4_1	0,609	0,682	0,708	0,711	0,681	0,791	0,596	0,667	0,652		1,000	0,708	0,720	0,711	0,583	0,681	0,638	0,622	0,583	0,708	0,604	0,766	0,556	0,627	0,636	0,636	0,650	0,511	0,400	0,391	0,372	0,293	0,409	0,500	0,449	0,449	0,468	0,565	0,552	0,582	0,632
4_5	0,667	0,651	0,766	0,727	0,696	0,810	0,609	0,681	0,667	1,000		0,766	0,776	0,727	0,638	0,652	0,609	0,591	0,596	0,638	0,654	0,783	0,566	0,600	0,605	0,651	0,667	0,522	0,408	0,356	0,381	0,300	0,465	0,553	0,500	0,458	0,522	0,578	0,561	0,556	0,607
4_8	0,609	0,636	0,750	0,800	0,638	0,698	0,681	0,750	0,739	0,708	0,766		0,680	0,578	0,542	0,511	0,511	0,578	0,625	0,667	0,566	0,638	0,667	0,706	0,727	0,773	0,650	0,468	0,480	0,435	0,419	0,293	0,409	0,417	0,449	0,490	0,681	0,609	0,621	0,655	0,702
5_4	0,708	0,565	0,640	0,723	0,653	0,667	0,694	0,680	0,667	0,720	0,776	0,680		1,000	0,520	0,571	0,735	0,553	0,680	0,640	0,764	0,653	0,607	0,642	0,652	0,565	0,476	0,408	0,538	0,375	0,400	0,326	0,348	0,480	0,510	0,392	0,531	0,625	0,600	0,632	0,678
5_7	0,698	0,634	0,622	0,667	0,727	0,650	0,773	0,667	0,698	0,711	0,727	0,578	1,000		0,533	0,636	0,818	0,571	0,667	0,578	0,680	0,636	0,588	0,583	0,683	0,585	0,541	0,455	0,511	0,419	0,450	0,368	0,390	0,489	0,565	0,391	0,591	0,698	0,618	0,654	0,667
6_3	0,609	0,591	0,583	0,578	0,553	0,605	0,553	0,583	0,565	0,583	0,638	0,542	0,520	0,533		0,723	0,596	0,533	0,542	0,625	0,528	0,723	0,519	0,588	0,545	0,591	0,650	0,468	0,520	0,391	0,372	0,390	0,545	0,500	0,694	0,490	0,511	0,478	0,552	0,545	0,456
6_6	0,489	0,605	0,553	0,545	0,565	0,619	0,565	0,553	0,578	0,681	0,652	0,511	0,571	0,636	0,723		0,565	0,591	0,468	0,638	0,538	0,609	0,566	0,560	0,512	0,558	0,513	0,522	0,408	0,311	0,429	0,400	0,465	0,426	0,583	0,458	0,478	0,622	0,561	0,556	0,500
7_3	0,711	0,605	0,511	0,591	0,565	0,524	0,696	0,638	0,667	0,638	0,609	0,511	0,735	0,818	0,596	0,565		0,727	0,681	0,596	0,615	0,565	0,528	0,600	0,651	0,512	0,564	0,435	0,490	0,444	0,333	0,300	0,326	0,426	0,583	0,458	0,565	0,667	0,702	0,741	0,714
7_8	0,605	0,634	0,533	0,571	0,500	0,600	0,591	0,578	0,558	0,622	0,591	0,578	0,553	0,571	0,533	0,591	0,727		0,533	0,622	0,520	0,545	0,471	0,542	0,488	0,585	0,649	0,409	0,426	0,419	0,300	0,263	0,341	0,356	0,478	0,522	0,455	0,558	0,618	0,615	0,593
8_2	0,652	0,591	0,542	0,667	0,596	0,605	0,596	0,667	0,609	0,583	0,596	0,625	0,680	0,667	0,542	0,468	0,681	0,533		0,708	0,566	0,553	0,630	0,627	0,636	0,636	0,500	0,511	0,520	0,348	0,372	0,293	0,318	0,625	0,531	0,490	0,553	0,522	0,586	0,655	0,632
10_6	0,565	0,727	0,625	0,711	0,638	0,744	0,638	0,667	0,652	0,708	0,638	0,667	0,640	0,578	0,625	0,638	0,596	0,622	0,708		0,566	0,681	0,519	0,745	0,727	0,727	0,650	0,511	0,520	0,391	0,419	0,341	0,409	0,583	0,531	0,490	0,553	0,609	0,586	0,618	0,632
11_2	0,667	0,490	0,604	0,640	0,692	0,583	0,577	0,642	0,588	0,604	0,654	0,566	0,764	0,680	0,528	0,538	0,615	0,520	0,566	0,566		0,692	0,576	0,643	0,571	0,531	0,533	0,500	0,509	0,431	0,500	0,478	0,449	0,491	0,556	0,333	0,500	0,667	0,540	0,533	0,613
11_4	0,667	0,605	0,723	0,727	0,696	0,762	0,565	0,596	0,578	0,766	0,783	0,638	0,653	0,636	0,723	0,609	0,565	0,545	0,553	0,681	0,692		0,491	0,600	0,605	0,651	0,667	0,522	0,490	0,356	0,381	0,350	0,512	0,511	0,583	0,417	0,522	0,578	0,526	0,519	0,571
13_3	0,538	0,520	0,556	0,667	0,528	0,490	0,604	0,630	0,654	0,556	0,566	0,667	0,607	0,588	0,519	0,566	0,528	0,471	0,630	0,519	0,576	0,491		0,667	0,600	0,640	0,478	0,491	0,536	0,462	0,531	0,468	0,480	0,519	0,618	0,473	0,604	0,577	0,719	0,721	0,762
13_4	0,653	0,511	0,706	0,792	0,680	0,565	0,760	0,863	0,816	0,627	0,600	0,706	0,642	0,583	0,588	0,560	0,600	0,542	0,627	0,745	0,643	0,600	0,667		0,809	0,681	0,605	0,520	0,566	0,531	0,522	0,409	0,426	0,510	0,538	0,462	0,640	0,694	0,689	0,655	0,700
14_1	0,667	0,600	0,727	0,878	0,744	0,615	0,791	0,818	0,857	0,636	0,605	0,727	0,652	0,683	0,545	0,512	0,651	0,488	0,636	0,727	0,571	0,605	0,600	0,809		0,800	0,722	0,512	0,522	0,571	0,513	0,378	0,450	0,455	0,533	0,489	0,744	0,714	0,704	0,745	0,717
15_2	0,571	0,650	0,682	0,780	0,651	0,667	0,698	0,682	0,714	0,636	0,651	0,773	0,565	0,585	0,591	0,558	0,512	0,585	0,636	0,727	0,531	0,651	0,640	0,681	0,800		0,833	0,465	0,522	0,476	0,462	0,432	0,550	0,455	0,578	0,578	0,698	0,571	0,667	0,706	0,679
15_4	0,579	0,667	0,650	0,703	0,615	0,629	0,615	0,650	0,684	0,650	0,667	0,650	0,476	0,541	0,650	0,513	0,564	0,649	0,500	0,650	0,533	0,667	0,478	0,605	0,722	0,833		0,513	0,429	0,474	0,400	0,364	0,500	0,400	0,537	0,488	0,564	0,526	0,520	0,553	0,531
16_2	0,444	0,512	0,468	0,545	0,565	0,429	0,478	0,596	0,578	0,511	0,522	0,468	0,408	0,455	0,468	0,522	0,435	0,409	0,511	0,511	0,500	0,522	0,491	0,520	0,512	0,465	0,513		0,490	0,533	0,524	0,450	0,512	0,596	0,542	0,458	0,478	0,489	0,491	0,519	0,500
17_5	0,500	0,435	0,440	0,553	0,449	0,444	0,612	0,560	0,542	0,400	0,408	0,480	0,538	0,511	0,520	0,408	0,490	0,426	0,520	0,520	0,509	0,490	0,536	0,566	0,522	0,522	0,429	0,490		0,667	0,622	0,698	0,652	0,560	0,667	0,471	0,490	0,417	0,567	0,526	0,508
18_1	0,364	0,381	0,435	0,512	0,489	0,341	0,533	0,478	0,500	0,391	0,356	0,435	0,375	0,419	0,391	0,311	0,444	0,419	0,348	0,391	0,431	0,356	0,462	0,531	0,571	0,476	0,474	0,533	0,667		0,780	0,718	0,667	0,435	0,553	0,426	0,489	0,409	0,571	0,566	0,509
18_2	0,390	0,359	0,465	0,500	0,524	0,421	0,524	0,512	0,537	0,372	0,381	0,419	0,400	0,450	0,372	0,429	0,333	0,300	0,372	0,419	0,500	0,381	0,531	0,522	0,513	0,462	0,400	0,524	0,622	0,780		0,778	0,718	0,558	0,591	0,318	0,524	0,537	0,491	0,480	0,462
18_3	0,308	0,324	0,341	0,368	0,400	0,333	0,450	0,390	0,410	0,293	0,300	0,293	0,326	0,368	0,390	0,400	0,300	0,263	0,293	0,341	0,478	0,350	0,468	0,409	0,378	0,432	0,364	0,450	0,698	0,718	0,778		0,757	0,488	0,619	0,381	0,450	0,410	0,471	0,417	0,400
18_5	0,381	0,350	0,455	0,439	0,465	0,462	0,512	0,455	0,476	0,409	0,465	0,409	0,348	0,390	0,545	0,465	0,326	0,341	0,318	0,409	0,449	0,512	0,480	0,426	0,450	0,550	0,500	0,512	0,652	0,667	0,718	0,757		0,545	0,711	0,400	0,512	0,429	0,556	0,510	0,453
19_1	0,522	0,500	0,500	0,489	0,511	0,558	0,511	0,583	0,522	0,500	0,553	0,417	0,480	0,489	0,500	0,426	0,426	0,356	0,625	0,583	0,491	0,511	0,519	0,510	0,455	0,455	0,400	0,596	0,560	0,435	0,558	0,488	0,545		0,653	0,449	0,511	0,478	0,517	0,509	0,491
20_2	0,596	0,533	0,490	0,522	0,500	0,545	0,625	0,571	0,596	0,449	0,500	0,449	0,510	0,565	0,694	0,583	0,583	0,478	0,531	0,531	0,556	0,583	0,618	0,538	0,533	0,578	0,537	0,542	0,667	0,553	0,591	0,619	0,711	0,653		0,480	0,625	0,511	0,644	0,643	0,552
23_1	0,383	0,489	0,408	0,478	0,417	0,409	0,542	0,531	0,553	0,449	0,458	0,490	0,392	0,391	0,490	0,458	0,458	0,522	0,490	0,490	0,333	0,417	0,473	0,462	0,489	0,578	0,488	0,458	0,471	0,426	0,318	0,381	0,400	0,449	0,480		0,417	0,426	0,610	0,607	0,517
26_1	0,667	0,558	0,638	0,682	0,609	0,571	0,652	0,681	0,711	0,468	0,522	0,681	0,531	0,591	0,511	0,478	0,565	0,455	0,553	0,553	0,500	0,522	0,604	0,640	0,744	0,698	0,564	0,478	0,490	0,489	0,524	0,450	0,512	0,511	0,625	0,417		0,756	0,667	0,704	0,643
26_L5.S.R.	0,636	0,571	0,652	0,698	0,667	0,537	0,711	0,739	0,773	0,565	0,578	0,609	0,625	0,698	0,478	0,622	0,667	0,558	0,522	0,609	0,667	0,578	0,577	0,694	0,714	0,571	0,526	0,489	0,417	0,409	0,537	0,410	0,429	0,478	0,511	0,426	0,756		0,679	0,642	0,655
27_2	0,607	0,519	0,586	0,655	0,561	0,491	0,737	0,724	0,750	0,552	0,561	0,621	0,600	0,618	0,552	0,561	0,702	0,618	0,586	0,586	0,540	0,526	0,719	0,689	0,704	0,667	0,520	0,491	0,567	0,571	0,491	0,471	0,556	0,517	0,644	0,610	0,667	0,679		0,892	0,836
27_4	0,604	0,549	0,545	0,692	0,630	0,520	0,704	0,691	0,717	0,582	0,556	0,655	0,632	0,654	0,545	0,556	0,741	0,615	0,655	0,618	0,533	0,519	0,721	0,655	0,745	0,706	0,553	0,519	0,526	0,566	0,480	0,417	0,510	0,509	0,643	0,607	0,704	0,642	0,892		0,875
28_2	0,582	0,528	0,632	0,704	0,607	0,538	0,714	0,702	0,727	0,632	0,607	0,702	0,678	0,667	0,456	0,500	0,714	0,593	0,632	0,632	0,613	0,571	0,762	0,700	0,717	0,679	0,531	0,500	0,508	0,509	0,462	0,400	0,453	0,491	0,552	0,517	0,643	0,655	0,836	0,875	

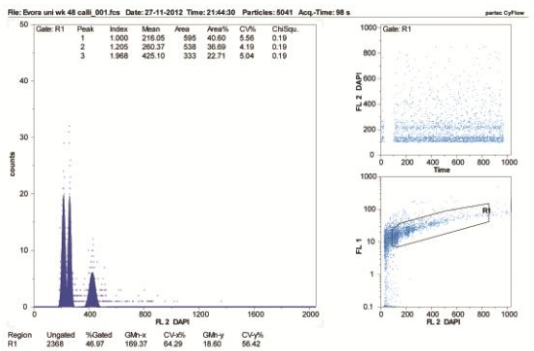
Appendix 5

Appendix 6 - **A** - *D. c*. 'Senta' (accession 2), **B** - *D. c.* 'Nantes normu' (accession 4), **C** - *D. c.* 'Lange rote stumpfe'(accession 8), **D** - *D. c. halophilus* (accession 18), **E** - *D. c. gummifer* (accession 19), **F** - *D. c. gadecaei* (accession 20), **G** - *D. c.* 'Rotin' (accession 26) and **H** - *D. c.* 'Nevis F1' (accession 27). **1** - Mericarp (A, B, C, D, E, F, G and H) germinated plant leaf control and **2** - Cell line *calli* (A2 - 2_4 - *D. c.* 'Senta'; B2 - 4_5 - *D. c.* 'Nantes normu'; C2 - 8_2 - *D. c.* 'Lange rote stumpfe'; D2 - 18_1 - *D. c. halophilus*; E2 - 19_1 - *D. c. gummifer*; F2 - 20_2 - *D. c. gadecaei*; G2 - 26_L5.S.R. - *D. c.* 'Rotin' and H2 - 27_2 - *D. c.* 'Nevis F1'). Flow cytometry analysis graphics (see sections 2.2.4 and 3.2.5).

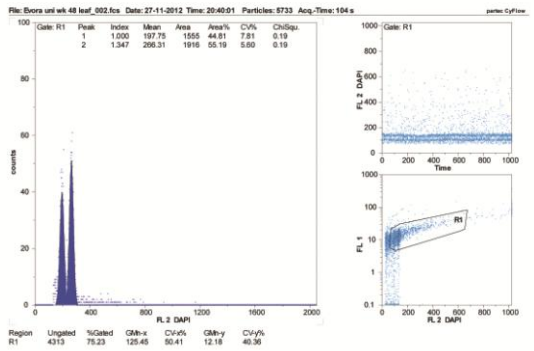
A1



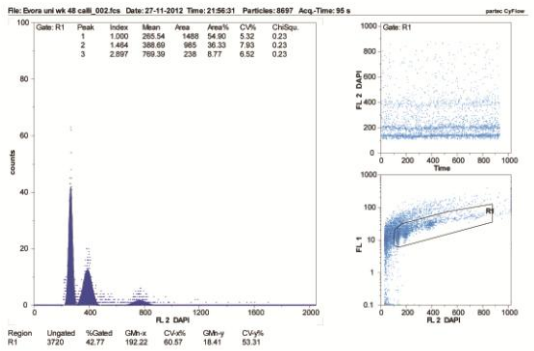
A2



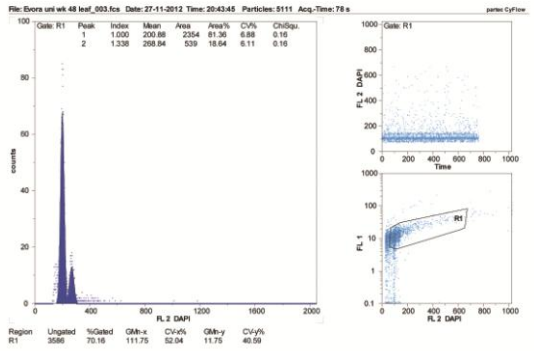
B1



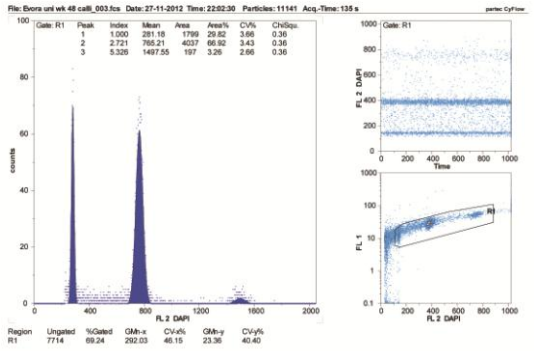
B2



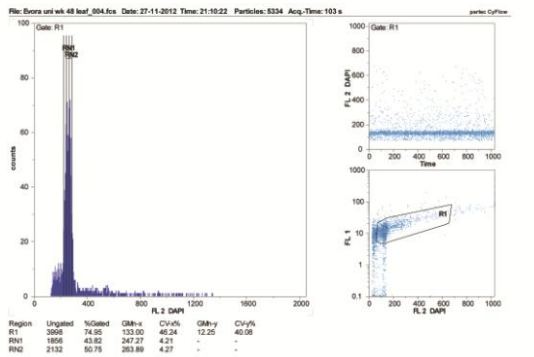
C1



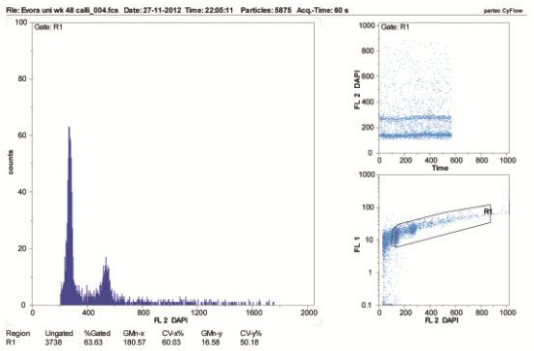
C2



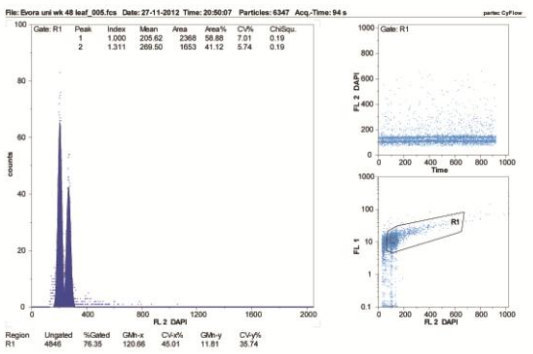
D1



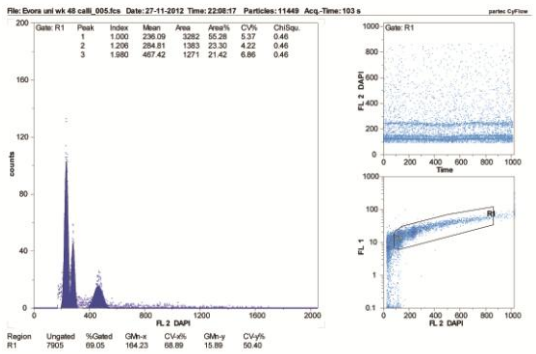
D2



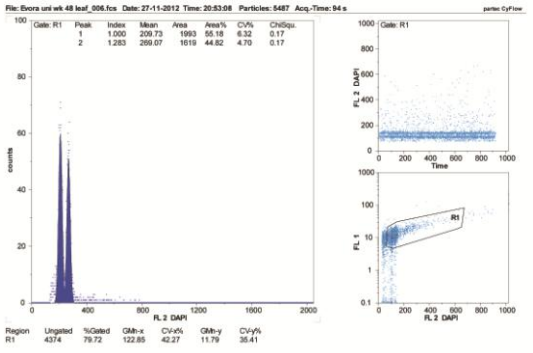
E1



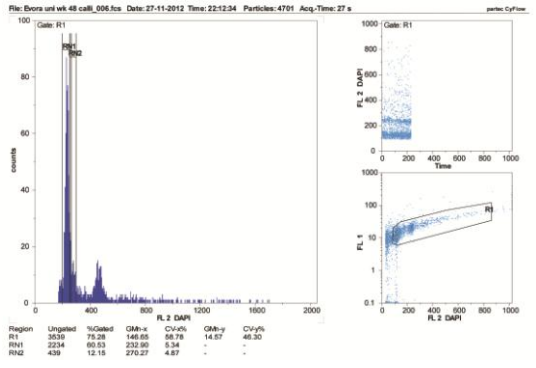
E2



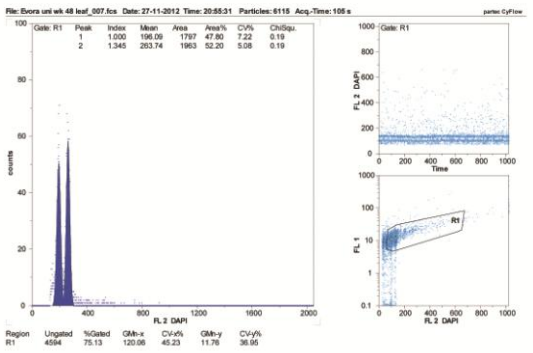
F1



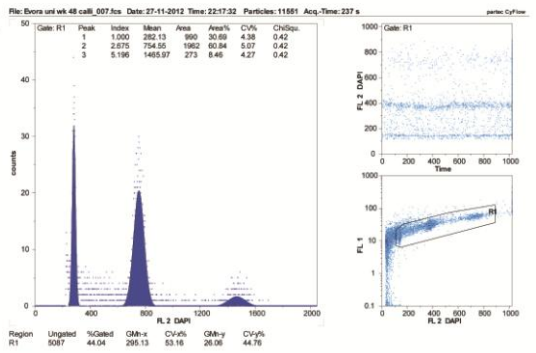
F2



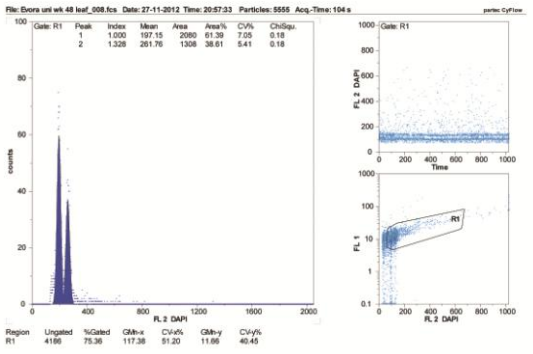
G1



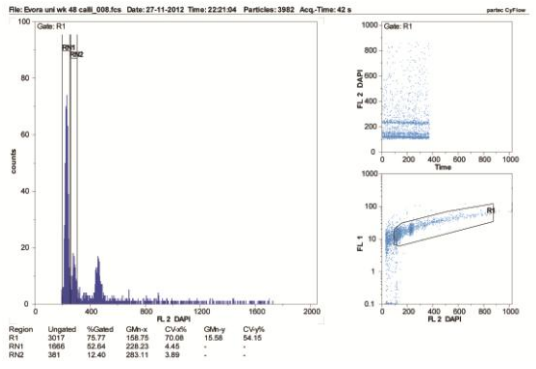
G2



H1



H2



Appendix 7 - **A** - *AOX1*, **B** - *AOX2a* and **C** - *AOX2b*. SNPs and InDels distribution and position across the aligned sequences of the identified *AOX* ORFs. Residue coding nucleotide triplets or InDel region's position numbering indicated in the table refers to aligned sequences (see section 2.2.6). The numbers below the table identifies individually the mutation for reference. The International Union of Pure and Applied Chemistry (IUPAC) amino acid one letter code and name is provided below as an additional table in the Appendix 7. **A. A.** - Amino acid; **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed; **Grey cells** - SNPs occurrence site. The italic nucleotide in the triplet indicates the difference among the sequences; **InDel** - Insertion and deletion; **nsSNP** - Non-synonymous SNP; **sSNP** - Synonymous SNP.

Additional table of Appendix 7 - IUPAC amino acid one letter code and name.

IUPAC code	Name
A	Alanine
B	Asparagine or Aspartic acid
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine
Z	Glutamine or Glutamic acid

A

	Cell line	Clone		Exon 1																														Exon 2								Exon 3
				sSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	sSNP	sSNP	sSNP	sSNP	nsSNP	InDel	nsSNP	nsSNP	nsSNP	InDel	nsSNP	sSNP	nsSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	nsSNP
				28-30	37-39	43-45	46-48	49-51	88-91	100-102	115-117	136-138	139-141	151-153	160-192	193-195	196-198	199-201	208-216	220-222	241-243	253-255	262-264	292-294	304-306	322-324	334-336	355-357	361-363	364-366	370-372	379-381	412-414	415-417	634-636	637-639	718-720	748-750	757-759	784-786	868-870	934-936
AOX1	2_4	50	DNA	GTC	TTG	ACG	GCC	GAC	AGC	CCG	CGG	TTG	GGT	CAG	G GA TCA TCA GCA GCA GCA GCA GGT GGT GGT GAT	AAT	AAA	CGT	GGT GGT AAT	GAA	GGG	GAA	AAG	AAC	CCC	GCG	ATC	GTC	ACC	ACG	TTG	CTT	CCT	ACT	CGC	GCT	GTC	CAC	ACT	GGC	ATG	CAA
			A. A.	V	L	T	A	D	S	P	R	L	G	Q	G S S A A A A G G G D	N	K	R	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
		51	DNA	GTC	TTG	ATG	GGT	GGC	GGC	CCG	CGG	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	GGT GGT AAT	GAA	GGA	GAA	AAG	AAC	CCA	GCT	ATC	GTT	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	L	M	G	G	G	P	R	L	G	Q	N	K	N	G	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
	4_5	1	DNA	GTC	TTG	ATG	GGT	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	GGT GGT AAT	GAA	GGA	GAA	AAG	AAC	CCA	GCT	ATC	GTT	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	L	M	G	G	S	P	R	L	G	Q	N	K	N	G	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
	8_2	4	DNA	GTC	TTC	ACG	GCC	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	-- --- ---	CAA	GGA	CAA	AAG	AAC	CCA	GCG	ATC	GTC	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	TAC	ACC	GGT	ATG	CAA
			A. A.	V	F	T	A	G	S	P	R	L	G	Q	N	K	N	G	-- --- ---	Q	G	Q	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	Y	T	G	M	Q
	18_1	1	DNA	GTG	TTG	ATG	GGT	GGC	GGC	CCA	CGG	CTG	GGG	GAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	AGT	GGT GGT AAT	GAA	GGA	GAA	AAA	AAT	CCC	GCT	ATA	GTC	ACC	ACG	TTG	CTG	CCG	ACT	CGT	GCT	GTC	CAC	ACT	GGC	ATA	CAA
			A. A.	V	L	M	G	G	G	P	R	L	G	E	N	K	N	S	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
	19_1	4	DNA	GTC	TTG	ATG	GGT	GGC	GGC	CCG	CGG	CTG	GGG	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	AGT	GGT GGT AAT	GAA	GGA	GAA	AAG	AAC	CCC	GCT	ATC	GTC	ACC	ACT	TTG	CTG	CCC	ACC	CGT	GCT	GTC	CAC	ACT	GGC	ATG	CGA
			A. A.	V	L	M	G	G	G	P	R	L	G	Q	N	K	N	S	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	R
	20_2	1	DNA	GTC	TTG	ATG	GGT	GGC	GGC	CCG	CGT	CTG	GGG	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	AGT	GGT GGT AAT	GAA	GGA	GAA	AAA	AAT	CCC	GCT	ATC	GTC	ACC	ACT	TTG	CTG	CCC	ACC	CGT	GCT	GTC	CAC	ACT	GGC	ATG	CGA
			A. A.	V	L	M	G	G	G	P	R	L	G	Q	N	K	N	S	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	R
	26_L5.S.R.	1	DNA	GTC	TTG	ATG	GGT	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	GGT GGT AAT	GAA	GGA	GAA	AAG	AAC	CCA	GCT	ATC	GTT	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	L	M	G	G	S	P	R	L	G	Q	N	K	N	G	G G N	E	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
	27_2	1	DNA	GTC	TTC	ACG	GCC	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	-- --- ---	CAA	GGA	GAA	AAG	AAC	CCA	GCG	ATC	GTC	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	F	T	A	G	S	P	R	L	G	Q	N	K	N	G	-- --- ---	Q	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
		47	DNA	GTC	TTC	ACG	GCC	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	-- --- ---	CAA	GGA	GAA	AAG	AAC	CCA	GCG	ATC	GTC	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	F	T	A	G	S	P	R	L	G	Q	N	K	N	G	-- --- ---	Q	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
		52	DNA	GTC	TTC	ACG	GCC	GGC	AGC	CCG	CGT	CTG	GGT	CAG	A-- --- --- --- --- --- --- --- --- --- -AT	AAG	AAT	GGT	-- --- ---	CAA	GGA	GAA	AAG	AAC	CCA	GCG	ATC	GTC	ACT	ACT	TTA	CTG	CCT	ACC	CGT	GCC	GTT	CAC	ACC	GGT	ATG	CAA
			A. A.	V	F	T	A	G	S	P	R	L	G	Q	N	K	N	G	-- --- ---	Q	G	E	K	N	P	A	I	V	T	T	L	L	P	T	R	A	V	H	T	G	M	Q
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39

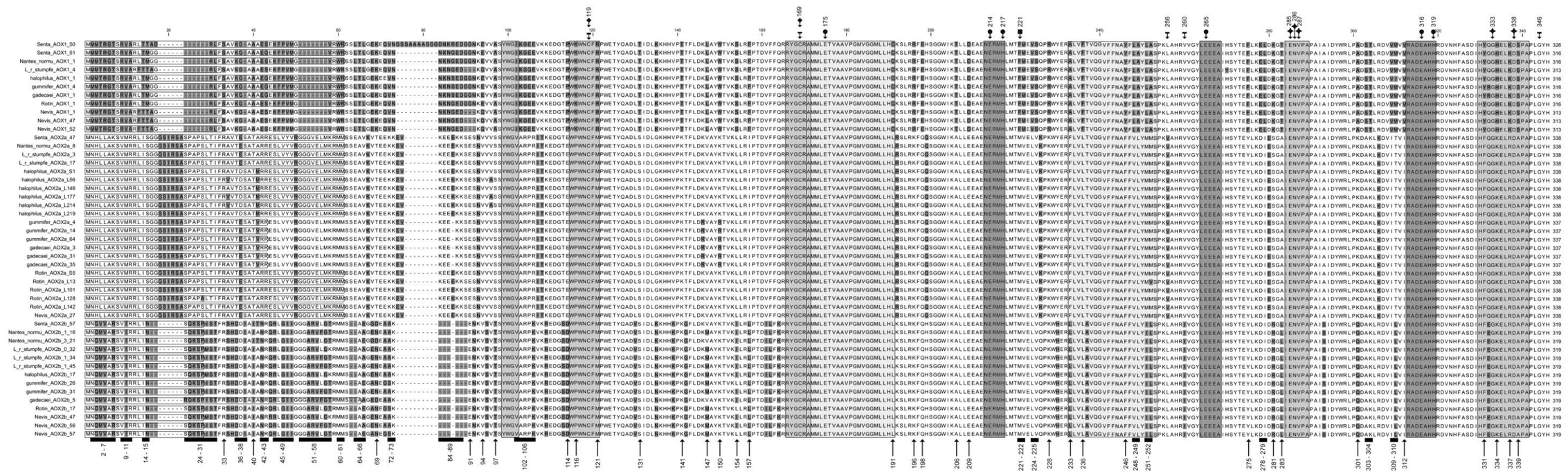
B

	Cell line	Clone		Exon 1												Exon 2				Exon 3							Exon 4
				nsSNP	nsSNP	sSNP	nsSNP	nsSNP	nsSNP	nsSNP / InDe	nsSNP	sSNP	sSNP	sSNP	sSNP	sSNP	nsSNP	nsSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	nsSNP	sSNP	sSNP
				52-54	82-84	85-87	100-102	109-111	121-123	235-237	262-264	280-282	298-300	310-312	331-333	376-378	412-414	424-426	460-462	526-528	541-543	592-594	664-666	685-687	730-732	844-846	973-975
AOX2a	2_4	47	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	4_5	8	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	8_2	3	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		17	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	18_1	S1	DNA	***A***GC	TCG	TTG	GCA	GA***T***	***A***TG	GAG	GTG	GT***A***	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TA***C***	ACC	ATG	CCT	GGG
			A. A.	**S**	**S**	**L**	**A**	**D**	**M**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L56	DNA	GGC	TCG	TTG	G***T***A	GA***T***	***A***TG	***A***AG	G***C***G	GTG	ACT	GG***A***	TG***C***	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**V**	**D**	**M**	**K**	**A**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L146	DNA	***A***GC	TCG	TTG	GCA	GA***T***	***A***TG	GAG	GTG	GT***A***	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**S**	**S**	**L**	**A**	**D**	**M**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L177	DNA	GGC	TCG	TTG	G***T***A	GA***T***	***A***TG	***A***AG	G***C***G	GTG	ACT	GG***A***	TG***C***	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**V**	**D**	**M**	**K**	**A**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L214	DNA	GGC	TCG	TTG	G***T***A	GA***T***	***A***TG	***A***AG	G***C***G	GTG	ACT	GG***A***	TG***C***	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**V**	**D**	**M**	**K**	**A**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L219	DNA	***A***GC	TCG	TTG	GCA	GA***T***	***A***TG	GAG	GTG	GT***A***	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**S**	**S**	**L**	**A**	**D**	**M**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	19_1	4	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		14	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		64	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	20_2	3	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		31	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		35	DNA	GGC	TCG	TTG	GCA	GAA	GTG	———	GTG	GTG	AC***G***	GGT	TG***C***	CT***C***	AG***G***	AG***G***	TT***T***	GTC	***T***TG	GC***A***	TAT	AC***T***	ATG	CC***G***	GG***A***
			A. A.	**G**	**S**	**L**	**A**	**E**	**V**	**—**	**V**	**V**	**T**	**G**	**C**	**L**	**R**	**R**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	26_L5.S.R	S5	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L13	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	***G***TG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**V**	**P**	**G**
		L101	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L128	DNA	GGC	***G***CG	***C***TG	GCA	GAA	G***C***G	***A***AG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**A**	**L**	**A**	**E**	**A**	**K**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
		L142	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GTC	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
	27_2	27	DNA	GGC	TCG	TTG	GCA	GAA	G***C***G	GAG	GTG	GTG	ACT	GGT	TGT	CTT	AAG	AAG	TTC	GT***T***	CTG	GCG	TAT	ACC	ATG	CCT	GGG
			A. A.	**G**	**S**	**L**	**A**	**E**	**A**	**E**	**V**	**V**	**T**	**G**	**C**	**L**	**K**	**K**	**F**	**V**	**L**	**A**	**Y**	**T**	**M**	**P**	**G**
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

C

				Exon 1														Exon 2					Exon 3														Exon 4
				nsSNP	nsSNP	sSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	nsSNP	sSNP	sSNP	sSNP	nsSNP	nsSNP	nsSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP	sSNP
	Cell line	Clone		10_12	46-48	49-51	61-63	97-99	106-108	142-144	148-150	166-168	172-174	175-177	181-183	184-186	247-249	265-267	304-306	319-321	322-324	358-360	379-381	427-429	442-444	493-495	511-513	664-666	709-711	757-759	781-783	802-804	832-834	853-855	856-858	862-864	910-912
AOX2b	2_4	57	**DNA**	***G***TG	AGC	CAG	ATT	***A***CG	CAG	TTG	ACA	GA***A***A	GG***G***G	***G***AT	GCG	GCG	GAG	CCT	GAC	CTG	AA***G***	***G***TG	ATT	CG***C***	GA***A***	AA***A***	CAG	TTA	TTG	GAT	CC***G***	GAT	CTG	ATT	CG***C***	GA***T***	TTC
			A. A.	**V**	**S**	**Q**	**I**	**T**	**Q**	**L**	**T**	**E**	**G**	**D**	**A**	**A**	**E**	**P**	**D**	**L**	**K**	**V**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	4_5	1_16	**DNA**	ATG	AGC	CA***A***	ATT	GCG	CAG	TT***T***	ACA	GCA	GAG	AAT	GCG	GCG	GAG	CC***G***	GAC	CTG	AAT	ATG	ATT	CG***C***	GA***A***	AAG	CAG	TTA	TT***A***	GAT	CC***G***	GAT	***T***TG	ATT	CG***C***	GA***T***	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**Q**	**F**	**T**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		3_21	**DNA**	***G***TG	AGC	CAG	ATT	***A***CG	CAG	TTG	ACA	GA***A***A	GG***G***G	***G***AT	GCG	GCG	GAG	CCT	GAC	CTG	AA***G***	***G***TG	ATT	CG***C***	GA***A***	AA***A***	CAG	TTA	TTG	GAT	CC***G***	GAT	CTG	ATT	CG***C***	GA***T***	TTC
			A. A.	**V**	**S**	**Q**	**I**	**T**	**Q**	**L**	**T**	**E**	**G**	**D**	**A**	**A**	**E**	**P**	**D**	**L**	**K**	**V**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	8_2	0_32	**DNA**	ATG	AGC	CAG	AT***G***	GCG	CG***G***	TTG	AA***A***	GCA	GAG	AAT	GCG	GCG	GAG	CCT	GA***T***	CTG	AAT	ATG	ATT	CG***T***	GA***G***	AAG	CAG	***C***TA	TTG	GAT	CCA	GAT	CTG	AT***A***	CGT	GAC	TT***T***
			A. A.	**M**	**S**	**Q**	**M**	**A**	**R**	**L**	**K**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		1_34	**DNA**	ATG	AGC	CA***A***	ATT	GCG	CAG	TT***T***	ACA	GCA	GAG	AAT	GCG	GCG	GAG	CC***G***	GAC	CTG	AAT	ATG	ATT	CG***C***	GA***A***	AAG	CAG	TTA	TTG	GA***C***	CCA	GAT	CTG	ATT	CG***C***	GAC	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**Q**	**F**	**T**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		1_45	**DNA**	ATG	AGC	CA***A***	ATT	GCG	CAG	TTG	ACA	GCA	GAG	AAT	GCG	GCG	GAG	CC***G***	GAC	CTG	AAT	ATG	ATT	CG***C***	GA***A***	AAG	CAG	TTA	TT***A***	GAT	CC***G***	GAT	***T***TG	ATT	CG***C***	GA***T***	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**Q**	**L**	**T**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	18_1	17	**DNA**	ATG	AGC	CAG	AT***G***	GCG	CG***G***	TTG	AA***A***	GCA	GAG	AAT	GCG	GCG	GAG	CCT	GA***T***	CTG	AAT	ATG	ATT	CG***T***	GA***G***	AAG	CAG	***C***TA	TTG	GAT	CCA	GAT	CTG	AT***A***	CGT	GAC	TT***T***
			A. A.	**M**	**S**	**Q**	**M**	**A**	**R**	**L**	**K**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	19_1	2_26	**DNA**	***G***TG	AGC	CAG	ATT	GCG	CAG	TTG	ACA	GA***A***	GG***G***	AAT	G***G***G	***T***CG	GA***A***	CCT	GAC	CTG	AAT	ATG	***C***TT	CG***T***	GA***G***	AA***A***	CA***A***	***C***TA	TTG	GAT	CCA	GAT	CTG	AT***A***	CGT	GAC	TTC
			A. A.	**V**	**S**	**Q**	**I**	**A**	**Q**	**L**	**T**	**E**	**G**	**N**	**G**	**S**	**E**	**P**	**D**	**L**	**N**	**M**	**L**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		2_31	**DNA**	ATG	AGC	CA***A***	ATT	GCG	CAG	TT***T***	ACA	GCA	GAG	AAT	GCG	GCG	GAG	CC***G***	GAC	CTG	AAT	ATG	ATT	CG***C***	GA***A***	AAG	CAG	TTA	TT***A***	GAT	CC***G***	GAT	***T***TG	ATT	CGT	GA***T***	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**Q**	**F**	**T**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	20_2	5	**DNA**	ATG	AG***A***	CAG	ATT	GCG	CAG	TT***T***	ACA	GCA	GAG	G***A***T	GCG	GCG	GAG	CC***G***	GAC	CT***A***	AAT	G***T***G	ATT	CG***T***	GA***G***	AA***A***	CAG	TTA	TTG	GAT	CCA	GA***C***	CTG	ATT	CGT	GAC	TTC
			A. A.	**M**	**R**	**Q**	**I**	**A**	**Q**	**F**	**T**	**A**	**E**	**D**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**V**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	26_L5.S.R.	17	**DNA**	***G***TG	AGC	CAG	AT***G***	GCG	CAG	TTG	ACA	GA***A***	GG***G***	AAT	G***G***G	***T***CG	GA***A***	CCT	GAC	CTG	AAT	ATG	***C***TT	CG***T***	GA***G***	AA***A***	CA***A***	***C***TA	TTG	GAT	CCA	GAT	CTG	AT***A***	CGT	GAC	TTC
			A. A.	**V**	**S**	**Q**	**M**	**A**	**Q**	**L**	**T**	**E**	**G**	**N**	**G**	**S**	**E**	**P**	**D**	**L**	**N**	**M**	**L**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
	27_2	47	**DNA**	ATG	AGC	CAG	ATT	GCG	CG***G***	TTG	AA***A***	GCA	GAG	AAT	GCG	GCG	GAG	CCT	GA***T***	CTG	AAT	ATG	ATT	CG***T***	GA***G***	AAG	CAG	***C***TA	TTG	GAT	CCA	GAT	CTG	AT***A***	CGT	GAC	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**R**	**L**	**K**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		56	**DNA**	ATG	AGC	CA***A***	ATT	GCG	CAG	TT***T***	ACA	GCA	GAG	AAT	GCG	GCG	GAG	CC***G***	GAC	CTG	AAT	ATG	ATT	CG***C***	GA***A***	AAG	CAG	TTA	TT***A***	GAT	CC***G***	GAT	***T***TG	ATT	CG***C***	GA***T***	TTC
			A. A.	**M**	**S**	**Q**	**I**	**A**	**Q**	**F**	**T**	**A**	**E**	**N**	**A**	**A**	**E**	**P**	**D**	**L**	**N**	**M**	**I**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
		57	**DNA**	***G***TG	AGC	CAG	ATT	GCG	CAG	TTG	ACA	GA***A***	GG***G***	AAT	G***G***G	***T***CG	GA***A***	CCT	GAC	CTG	AAT	ATG	***C***TT	CG***T***	GA***G***	AA***A***	CA***A***	***C***TA	TTG	GAT	CCA	GAT	***T***TG	AT***A***	CGT	GAC	TTC
			A. A.	**V**	**S**	**Q**	**I**	**A**	**Q**	**L**	**T**	**E**	**A**	**N**	**G**	**S**	**E**	**P**	**D**	**L**	**N**	**M**	**L**	**R**	**E**	**K**	**Q**	**L**	**L**	**D**	**P**	**D**	**L**	**I**	**R**	**D**	**F**
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34

Appendix 8 - Multiple alignment of the 47 AOX translated proteins identified in the eight selected cell lines. Grey residues indicate the differences amongst the sequences. The "-" signal represent a gap in the alignment. Annotated sites below the alignment represent the residues indicated by the SH method as AOX1 and AOX2 putatively of functional relevance (see section 3.4.1). Annotated sites above alignment and boxes represent the residues with known functional relevance, which had been published and are used at the UniProt database (www.uniprot.org) as reference for AOX characterization. The International Union of Pure and Applied Chemistry (IUPAC) amino acid one letter code and name is provided as an additional table in the Appendix 7. **Arrow boxes** - Targeting peptide cleavage site sequences (see section 3.4.1); **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code (number after the accession name) identifying the bacterial clone and from which the sequence code was attributed; **Dark-grey boxes** - Highly conserved regions in AOX sequences (Berthold *et al.*, 2000); **Diamond arrows** - Sites of the two conserved cysteine residues (Cys_I and Cys_{II}) indicated by Umbach and Siedow (1993) and later revised by Rhoads *et al.* (1998), as being involved in the formation of the disulfide bond; **Light-grey boxes** - Reference transmembrane helical regions (modified from Berthold *et al.*, 2000; Saisho *et al.*, 2001; Heazlewood *et al.*, 2004); **Mid-grey boxes** - Structural elements proposed to influence AOX regulatory behavior (Crichton *et al.*, 2005); **Rectangular arrows** - Residues already tested by direct site mutagenesis, which had an effect on protein biological properties (Berthold, 1998; Rhoads *et al.*, 1998; Djajanegara *et al.*, 1999; Umbach *et al.*, 2002); **Round arrows** - Iron cofactor binding residues (Siedow *et al.*, 1995; Andersson and Nordlund, 1999), revised by Berthold *et al.* (2000); **Square arrow** - Residue indicated by Frederico *et al.* (1999b) for the identification of both AOX subfamilies in higher plants; **Star arrows** - Residues potentially involved in regulation of AOX activity (Crichton *et al.*, 2005).



Appendix 9 - **A** - *AOX1* ORF, **B** - *AOX2a* ORF, **C** - *AOX2b* ORF, **D** - *AOX1* intron1, **E** - *AOX2a* intron 1 and **F** - *AOX2b* intron 1. Regulatory elements enrichment assessment in the *AOX* ORF and intron 1 unaligned nucleotide sequences identified in the eight cell lines with extreme SE efficiency phenotypes using the software Nsite-PL and ScanWM-P (see section 2.2.7, 3.4.2 and 3.4.3). The assessment was performed individually for each gene sequence (Tables A to F) using both softwares, as indicated separately on each table. The position numbering indicated next to the regulatory element refers to the unaligned position and the orientation of the regulatory element (forward or reverse strand). The displayed position of the regulatory element in the table refers to the aligned position (see section 2.2.6). The number inside parentheses near the elements indicate the number of different matches found (usually from a different species/genes). The number below each individual table was used to identify individually the position as reference. Additional tables with the *cis*-acting regulatory element and binding factor name/identification are provided below as part of the Appendix 9. **BF** - Binding factor (when absent, consider as "unknown"); **Cell line** - 2_4 - *D. c.* 'Senta'; 4_5 - *D. c.* 'Nantes normu'; 8_2 - *D. c.* 'Lange rote stumpfe'; 18_1 - *D. c. halophilus*; 19_1 - *D. c. gummifer*; 20_2 - *D. c. gadecaei*; 26_L5.S.R. - *D. c.* 'Rotin'; 27_2 - *D. c.* 'Nevis F1'; **Clone** - Code identifying the bacterial clone and from which the sequence code was attributed; **Grey lines** - Sequences from non-embryogenic cell lines; **RE** - Regulatory element (when absent, consider as "unknown").

Additional tables of Appendix 9 - Name/identification of the *cis*-acting regulatory element (RE) and binding factor (BF) displayed on the Appendix 9 tables (A, B, C, D, E and F).

RE	Name / Identification
5A W	5A element W from *Arabidopsis* LHY gene
- 300	−300 bp element from rice glutelin Gt3 gene
A1	HVA1 motif from barley GCCGAC gene
ABRE1/2/3/4/A/B	ABA response elements
AC1	AC 1 element
AC-I	AC-I element
AE 2	AE box 2
AT1	AT1 motif element
AGCC	AGCC motif element
Alfin1	Alfalfa salt tolerance element
AGL2	Agamous-like binding element
Amylase	Amylase element
AT1	AT-1 motif element
ATCATC	ATCATC motif element
ATF	Activating transcription factor element
AuxRE	Auxin response element
B1	Aux28 G region of soybean Aux28 B gene
BN	BN element from Catharanthus Str gene
Box 1	Box 1 element
Box 2	Box 2 element
Box A	Box A element
Box A2	Box A2 element from oat alpha-Amy2 gene
Box I	Box I element
Box II	Box II element
CE1/3	Coupling elements
Box III	Box III element (H box)
Box L	Box L promoter element
Box V	Box V from *Arabidopsis* S1 gene
bZIP	Basic leucine-zipper domain element
C1	C1 box element
C2a	Heat stress transcription factor C2a element
CAMTA1/5	Calmodulin-binding transcription activators
CARE D/H	CAACTC regulatory elements
CAT	Chloramphenicol acetyl transferase element
CCA1	Circadian clock associated 1 element
CCAAT	CCAAT motif element
CCAAAT	CCAAAT motif element
CCGTCC	CCGTCC motif element
CCGTTA	CCGTTA motif element
CGACG	CGACG motif element
CM2	Conserved DNA motif 2 element
Cp	Cysteine protease element
CRE 33-36/39-42/45-48	cis-regulatory elements
C-rich	C-rich element
CT-rich	CT-rich element
Distal	Distal-motif from *Arabidpsis* COX5b gene
DLEC2A	Phytohemagglutinin element
DRE1	Dehydration-responsive element 1
DOF1	DNA-binding with one finger element
DREmut1	Dehydration-responsive element mut 1
E2/4/5	E-core site elements
EE	Evening element or Timing of CAB expression 1 (TOC1)
EIN3	Ethylene insensitive3
Element 1	Element 1 from soybean Ibc3 gene
ElRE	Elicitor responsive element
MNF1	Maize nuclear factor 1 element
EM1	Early methionine 1 element
Em1b	Early methionine 1b element
ERE 2	Ethylene-responsive element 2
FHY3/FAR1	Far-red elongated hypocotyl3/Far-red impaired response 1
FLS2 EIN3/EIL1	Flagellin-sensing2 ethylene insensitive3/ EIN3-like1
Fp6/II	Fp6/II motif from soybean CHS8 gene
Fp12/III	Fp12/III motif from soybean CHS8 gene

BF	Name / Identification
ABF1/2/3/4	ABRE binding factors
ABI3/4/5	ABA-insensitive factors
AEF	Adult enhancer factor
AGAMOUS	MADS domain transcription factor agamous
AGL2	Agamous-like factor 2
aleurone	Aleurone layers nuclear protein extracts
Alfin1	Alfalfa salt tolerance factor
AP1	Activator protein 1
AREB1	ABRE-binding factor 1
AT1	AT-1 motif factor
BBBF	Box B binding factor
BEL5	Potato YUCCA gene BEL5 domain factor
bHLH122	Basic-helix-loop-helix factor 122
BLZ1	Barley leucine zipper 1
BPC1	Basic pentacysteine 1
Bzip21	Basic leucine-zipper-transcription factor 21
C/EBP	CCAAT/enhancer binding protein
CAMTA1/3/5	Calmodulin-binding transcription activators
CBF1/2	C-repeat binding factors
CCA1	Circadian clock associated 1 factor
CIB1/2/3/4/5	Cryptochrome-interacting basic-helix-loop-helix factors
Cp	Cysteine protease factor
CRR1	Copper response regulator
DEL65	DEL65 basic-helix-loop-helix
DF1	Trihelix DNA-binding domain DF-1
DOF3	DNA-binding with one finger factor 3
DPBF1/2	Dc3 promoter-binding factors
DREB3	Dehydration-responsive element-binding factor 3
EIN3	Ethylene-insensitive 3
EIN3/EIL1	Ethylene-insensitive 3/ Ethylene-insensitive 3-like1
EmBP1	Early methionine-binding protein 1
ERF1	Ethylene response factor 1
ESE1	Ethylene and salt inducible ERF genes factor 1
FHY3/FAR1	Far-red elongated hypocotyl3/Far-red impaired response 1
GAMYB	GA regulated MYB
GBF1	G-box binding factor 1
GBP	GAGA binding protein
GL3	Glabra 3
GT1/2	Trihelix transcription factors GT
H21	Barley H21 element from SyntheticOLIGOs gene
HsfA2	Heat stress transcription factor A
IPA1	Ideal plant architecture factor
KN1/KIP	knotted1/ knotted interacting protein
LFU	LFU motif from *Arabidopsis* agamous gene
LFY	Leafy factor
LHY	Late elongated hypocotyl
MADS	MCM1/Agamous/Deficiens/SRF-domain proteins factor
MAT2	Maturation regulation (ROM2)
MINI3	Miniseed 3 element factor
MNF1/B1a/1b	Maize nuclear factors
MYB2/3/4/20/21/61	MYB protein factors
MYC2/3/4	MYC protein factors
nodule factor	Nodule specific factor
NSP1	Nodulation signaling pathway 1
nuclear protein	Nuclear extract protein
O2	Opaque-2
OCSTF	Octopine synthase trnscription factor
p33TCP20	*Arabidopsis* TCP20 gene
PacC	*Aspergillus* pH-responsive
PAN	Perianthia factor
PBF-1	Prolamin-box binding factor 1
PCF1/2	Proliferating cell nuclear antigen factors
PG1	Paralogous group 1 factor

Additional tables of Appendix 9 - *Continued*

RE	Name / Identification
G	G box motif element
GA1/2/5	GA motif elements
GAAATA	GAAATA motif element
GAMYB	GA regulated MYB element
GARE 1	Gibberellin response element
GATA	GATA motif element
GBF1	G-box binding factor 1 element
GCCAAG	GCCAAG motif element
GSN	See hor1 element
GT1/2/K	Trihelix transcription factor GT element
GTAC1/2/3	GTAC motif elements
H21	Barley H21 element from SyntheticOLIGOs gene
16 KK	Tomato rbcS3A gene 16 kk motif element
hor1/2	B-hordein elements
HSE4	Heat shock element 4
I-box	I box element
ICEr2	Inducer of CBF expression region 1
Inr	Initiator motif element
KN1/KIP	Knotted1/ knotted interacting protein element
LBD	LOB domain element
LBS/WBS1	LFY/WUS1 binding sites element
LFY	Leafy element
LRE-TATA AA2	Tomato light-regulatory element -TATA AA2
MSA	Mitosis-specific activator element
MYC2/3/4/3	MYC protein elements
NDE1	*Nde*I restriction endonuclease site element
NF-kB-box	Nuclear factor kappa B box
Non	Wheat H3 gene unknown element
O2d	Opaque-2d binding site element
ocs	Octopine synthase element
P1/2	P boxes (Prolamin-boxes)
p33TCP20	*Arabidopsis* TCP20 gene element
PacC BS	*Aspergillus* pH-responsive element
PAN	*Perianthia* element
PAT2	Potato Patatin 21 gene element 2
PB	Prolamin box element
PRD 2	Positive regulatory domain 2 element
PRE1	Photoreceptor regulatory element 1
PY	Pyrimidine box element
R	R motif element
REbeta	Phytochrome beta regulatory element
RIN	Ripening inhibitor element
RTBP1	Regulator of transcription factor IID 1
RY	RY/Sph element
SA/MJ-RE	SA/MJ regulatory element from *Agrobacterium* NOS gene
SEF3/3-2	Soybean embryo factor elements
Site II	Site II element
SMRE1	Secondary wall MYB-responsive element
SP8b	Sweet potato sporamin and beta-amylase gene 8b element
STRE	Stress response promoter element
T/G	T/G box motif element
TAGTCAAC	TAGTCAAC motif element
TATCCAT/C	TATCCAT/C motif element
TCGTGT	TCGTGT motif element
TGA1	TGA element 1
TGGGCC/T	TGGGCC/T motif element
TGTCACA	TGTCACA motif element
TL1	Translocon 1 element
UV	Ultraviolet element
W1/2/3	W box elements
Wi	MINI3 Wi box element
WOX11	WUS homeobox 11 element
WRKY53	WRKY element 53
WT	WT box element
WUS2	Wuschel 2
ZPT2-2	TFIIIA-type zinc finger motif from petunia

BF	Name / Identification
PHR1	Phosphatase regulatory factor
RAV	Related to ABI3/ Viviparous 1 (VP1)
R-GATA	GATA class R
RIN	Ripening inhibitor factor
RITA1	Rice bZIP transcriptional activator 1
ROM1/2	Repressors of maturation
root factor	Root-specific nuclear factor
RTCS	Rootless concerning crown and seminal roots
RVE1/2/3/4/7/8	Reveille binding factors
SBP	Squamosa-promoter binding protein factor
seed protein	Seed-specific protein factor
SEF3	Soybean embryo factor 3
SGBF1/2	Soybean G-box binding factors
SPF1	Sweet potato SP8a and SP8b DNA-binding proteins factor
SPF1	Sweet potato factor 1
TAF1	TATA-Box binding protein associated factor 1
TB1	Teosinte branched 1 factor
TBF1	TATA-binding protein factor 1
TCP20	*Teosinte*/*cycloidea*/PCF 20 factor
TFHP1	TFHP-1 protein factor
TGA1/5/6	TGA transcription factors
TLP11	TATA-binding protein-like protein 11 factor
WOX11	WUS homeobox 11
WPBF	Wheat prolamin-box binding factor
WRKY1/2/3/4/33/70	WRKY domain transcription factors
WUS	Wuschel (consensus)
ZPT2-2	TFIIIA-type zinc finger from petunia

A *AOX1*

Nsite-PL

Cell line	Clone	Lenght (bp)	RE: distal / BF: ESE1 (2)	RE: LRE-TATA AA2	RE: E2 / BF: DPBF1/2 RE: ABRE2 / BF: ABF1/2/3/4	RE: W / BF: WRKY, RAV	RE: / BF: BEL5	RE: MNF1 /BF: MNF1/B1a/1b	RE: SP8b /BF: SPF1
2_4	50	981		188-GTGATAA -194	197-AACGTGAA-204				
	51	951				290-AGGCTGACT-298	294-TGACTTGAC-302	605-GTGCCCTT-612	756-TACTATT-762
4_5	1	951				290-AGGCTGACT-298	294-TGACTTGAC-302	605-GTGCCCTT-612	756-TACTATT-762
8_2	4	942	187-CCACTTG-181 (2)					596-GTGCCCTT-603	756-TACTATT-762
18_1	1	951				290-AGGCTGACT-298	294-TGACTTGAC-302		
19_1	4	951				290-AGGCTGACT-298	294-TGACTTGAC-302		
20_2	1	951				290-AGGCTGACT-298	294-TGACTTGAC-302		
26_L5.S.R.	1	951				290-AGGCTGACT-298	294-TGACTTGAC-302	605-GTGCCCTT-612	756-TACTATT-762
27_2	1	942	187-CCACTTG-181 (2)					596-GTGCCCTT-603	756-TACTATT-762
	47	942	187-CCACTTG-181 (2)					596-GTGCCCTT-603	756-TACTATT-762
	52	942	187-CCACTTG-181 (2)					596-GTGCCCTT-603	756-TACTATT-762
			1	2	3	4	5	6	7

ScanWM-P

Cell line	Clone	Lenght (bp)	RE: A1 /BF: CBF1	RE: A1 /BF: CBF2	RE: TGA1	RE: R RE: A1/BF: CBF1	RE: CGACG	RE: A1 /BF: CBF1	RE: Non	RE: SA/MJ-RE	RE: bZIP	RE: Non	RE: Box L	RE: ATCATC	RE: BOX III /BF: GT1	RE: Alfin1 /BF: Alfin1
2_4	50	981	33-GGCGAC-28	35-GGTTGACGACGGCC-48	38-TGACGACG-45	39-GACGAC-44	41-CGACG-45	46-GCCGAC-51		131-TGACGT-136	136-ACGTCA-131	142-CACCCAACG-134	145-TCTCACCCAAC-135	162-ATCATC-167		
	51	951	33-GGCGAC-28		38-TGACGACG-45											
4_5	1	951	33-GGCGAC-28		38-TGACGACG-45				123-CCTCCAACG-115							
8_2	4	942	33-GGCGAC-28			39-CACGAC-44 (2)	41-CGACG-45		123-CCTCCAACG-115							181-CAAGTGGTGGCT-192
18_1	1	951			38-TGACGACG-45										180-ACCATCTTCACT-169	
19_1	4	951	33-GGCGAC-28		38-TGACGACG-45										180-ACCATCTTCACT-169	
20_2	1	951	33-GGCGAC-28		38-TGACGACG-45				123-CCTCCAACG-115						180-ACCATCTTCACT-169	
26_L5.S.R.	1	951	33-GGCGAC-28		38-TGACGACG-45				123-CCTCCAACG-115							
27_2	1	942	33-GGCGAC-28			39-CACGAC-44 (2)	41-CGACG-45		123-CCTCCAACG-115							181-CAAGTGGTGGCT-192
	47	942	33-GGCGAC-28			39-CACGAC-44 (2)	41-CGACG-45		123-CCTCCAACG-115							181-CAAGTGGTGGCT-192
	52	942	33-GGCGAC-28			39-CACGAC-44 (2)	41-CGACG-45		123-CCTCCAACG-115							181-CAAGTGGTGGCT-192
			1	2	3	4	5	6	7	8	9	10	11	12	13	14

B *AOX2a*

Nsite-PL

Cell line	Clone	Length (bp)	RE: CE1 / BF: ABI4	RE: TL1 / BF: TBF1	RE: Alfin1	RE: DRE1	RE: GT1	RE: RIN /BF: RIN
2_4	47	1017	54-GCCACCG-48					
4_5	8	1017	54-GCCACCG-48					
8_2	3	1017	54-GCCACCG-48					
	17	1017	54-GCCACCG-48					
18_1	S1	1017						
	L56	1017	54-GCCACCG-48	234-GAAGAAGAA-242	285-CGCCACT-279			
	L146	1017						
	L177	1017	54-GCCACCG-48	234-GAAGAAGAA-242	285-CGCCACT-279			
	L214	1017	54-GCCACCG-48	234-GAAGAAGAA-242	285-CGCCACT-279			
	L219	1017						
19_1	4	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
	14	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
	64	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
20_2	3	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
	31	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
	35	1014	54-GCCACCG-48			378-ACCGAGA-372	587-AGGCATT-593	
26_L5.S.R.	S5	1017	54-GCCACCG-48					
	L13	1017	54-GCCACCG-48					730-CCATATAAAG-721
	L101	1017	54-GCCACCG-48					
	L128	1017	54-GCCACCG-48	234-GAAGAAGAA-242				
	L142	1017	54-GCCACCG-48					
27_2	27	1017	54-GCCACCG-48					
			1	2	3	4	5	6

ScanWM-P

Cell line	Clone	Length (bp)	RE: A1 /BF: CBF1
2_4	47	1017	86-AACGAC-81
4_5	8	1017	86-AACGAC-81
8_2	3	1017	86-AACGAC-81
	17	1017	86-AACGAC-81
18_1	S1	1017	86-AACGAC-81
	L56	1017	86-AACGAC-81
	L146	1017	86-AACGAC-81
	L177	1017	86-AACGAC-81
	L214	1017	86-AACGAC-81
	L219	1017	86-AACGAC-81
19_1	4	1014	86-AACGAC-81
	14	1014	86-AACGAC-81
	64	1014	86-AACGAC-81
20_2	3	1014	86-AACGAC-81
	31	1014	86-AACGAC-81
	35	1014	86-AACGAC-81
26_L5.S.R.	S5	1017	86-AACGAC-81
	L13	1017	86-AACGAC-81
	L101	1017	86-AACGAC-81
	L128	1017	
	L142	1017	86-AACGAC-81
27_2	27	1017	86-AACGAC-81
			1

C *AOX2b*

Nsite-PL

Cell line	Clone	Length (bp)	RE: LBS/WBS1 /BF: LFU, WUS	RE: E2 /BF: DPBF1/2	RE: A1 /BF: CBF1	RE: G	RE: ABRE1	RE: ABRE	RE: ABRE2
2_4	57	960		13-CCACTTG-7		302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
4_5	1_16	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
	3_21	960		13-CCACTTG-7		302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
8_2	0_32	960	2-TGAATCAAATGG-13						
	1_34	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
	1_45	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
18_1	17	960	2-TGAATCAAATGG-13						
19_1	2_26	960		13-CCACTTG-7	188-TTCGAC-183	302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
	2_31	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
20_2	5	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
26_L5.S.R.	17	960		13-CCACTTG-7	188-TTCGAC-183	302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
27_2	47	960	2-TGAATCAAATGG-13						
	56	960	2-TGAATCAAATGG-13			302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
	57	960		13-CCACTTG-7	188-TTCGAC-183	302-CAGACGTGTC-311	304-GACGTG-309 (2)	305-ACGTG-309	309-CACGTC-304
			1	2	3	4	5	6	7

ScanWM-P

Cell line	Clone	Length (bp)	RE: Em1b /BF: EmBP1	RE: DLEC2A /BF: MAT2	RE: MSA	RE: ABRE3	RE: ABRE4	RE: ABRE	RE: Em1b /BF: EmBP1
2_4	57	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
4_5	1_16	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
	3_21	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
8_2	0_32	960			443-AGACCGTGG-451	854-TACGTG-859 (2)	859-CACGTA-854	855-ACGTG-859	855-ACGTGCAGA-863
	1_34	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
	1_45	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
18_1	17	960			443-AGACCGTGG-451	854-TACGTG-859 (2)	859-CACGTA-854	855-ACGTG-859	855-ACGTGCAGA-863
19_1	2_26	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300	443-AGACCGTGG-451	854-TACGTG-859			855-ACGTGCAGA-863
	2_31	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
20_2	5	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300	443-AGACCGTGG-451	854-TACGTG-859			855-ACGTGCAGA-863
26_L5.S.R.	17	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300	443-AGACCGTGG-451	854-TACGTG-859	859-CACGTA-854	855-ACGTG-859	
27_2	47	960			443-AGACCGTGG-451	854-TACGTG-859 (2)	859-CACGTA-854	855-ACGTG-859	855-ACGTGCAGA-863
	56	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300					
	57	960	305-ACGTGTCCA-313	311-GACACGTCTGCT-300	443-AGACCGTGG-451	854-TACGTG-859 (2)	859-CACGTA-854	855-ACGTG-859	855-ACGTGCAGA-863
			8	9	10	11	12	13	14

A *AOX1*

Nsite-PL

Cell line	Clone	Lenght (bp)
2_4	50	981
	51	951
4_5	1	951
8_2	4	942
18_1	1	951
19_1	4	951
20_2	1	951
26_L5.S.R.	1	951
27_2	1	942
	47	942
	52	942

ScanWM-P

Cell line	Clone	Lenght (bp)	RE: Alfin1 /BF: Alfin1	RE: E2 /BF: DPBF1/2	RE: ABRE	RE: Em1b /BF: EmBP1	RE: EIRE /BF: WRKY1/2/3	RE: W /BF: WRKY	RE: W /BF: WRKY	RE: TGA1	RE: C-rich	RE: ABRE	RE: G / BF: RITA1, bZIP	RE: C-rich	RE: A1 /BF: CBF1	RE: W /BF: WRKY	RE: W /BF: WRKY
2_4	50	981	179-CAGGTGGTGGTG-190		198-ACGTG-202	198-ACGTGAAGA-206				329-TGACGATC-336	357-CCCCACCAC-365	367-ACGTG-363	368-AACGTG-363		723-AACGAC-718		841-GCTGACT-847
	51	951							292-GCTGACT-298	329-TGACGATC-336							
4_5	1	951							292-GCTGACT-298	329-TGACGATC-336							
8_2	4	942		187-CCACTTG-181			216-TTGACC-211	217-CTTGACC-211		329-TGACGATC-336							
18_1	1	951							292-GCTGACT-298		357-CCCCACCAC-365	367-ACGTG-363	368-AACGTG-363		723-AACGAC-718	840-TATGACT-834	
19_1	4	951							292-GCTGACT-298	329-TGACGATC-336	357-CCCCACCAC-365			381-CCCCACCGA-389	723-AACGAC-718		
20_2	1	951							292-GCTGACT-298	329-TGACGATC-336	357-CCCCACCAC-365			381-CCCCACCGA-389	723-AACGAC-718		
26_L5.S.R.	1	951							292-GCTGACT-298	329-TGACGATC-336							
27_2	1	942		187-CCACTTG-181						329-TGACGATC-336							841-GCTGACT-847
	47	942		187-CCACTTG-181						329-TGACGATC-336							
	52	942		187-CCACTTG-181						329-TGACGATC-336							
			15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

B *AOX2a*

Nsite-PL

Cell line	Clone	Length (bp)
2_4	47	1017
4_5	8	1017
8_2	3	1017
	17	1017
18_1	S1	1017
	L56	1017
	L146	1017
	L177	1017
	L214	1017
	L219	1017
19_1	4	1014
	14	1014
	64	1014
20_2	3	1014
	31	1014
	35	1014
26_L5.S.R.	S5	1017
	L13	1017
	L101	1017
	L128	1017
	L142	1017
27_2	27	1017

ScanWM-P

Cell line	Clone	Length (bp)
2_4	47	1017
4_5	8	1017
8_2	3	1017
	17	1017
18_1	S1	1017
	L56	1017
	L146	1017
	L177	1017
	L214	1017
	L219	1017
19_1	4	1014
	14	1014
	64	1014
20_2	3	1014
	31	1014
	35	1014
26_L5.S.R.	S5	1017
	L13	1017
	L101	1017
	L128	1017
	L142	1017
27_2	27	1017

C *AOX2b*

Nsite-PL

Cell line	Clone	Length (bp)
2_4	57	960
4_5	1_16	960
	3_21	960
8_2	0_32	960
	1_34	960
	1_45	960
18_1	17	960
19_1	2_26	960
	2_31	960
20_2	5	960
26_L5.S.R.	17	960
27_2	47	960
	56	960
	57	960

ScanWM-P

Cell line	Clone	Length (bp)
2_4	57	960
4_5	1_16	960
	3_21	960
8_2	0_32	960
	1_34	960
	1_45	960
18_1	17	960
19_1	2_26	960
	2_31	960
20_2	5	960
26_L5.S.R.	17	960
27_2	47	960
	56	960
	57	960

D — *AOX1 intron 1* — Nsite-PL / ScanWM-P

	Cell line	Clone	Lenght (bp)	RE: CCAAT /BF: C/EBP RE: STRE	RE: Cp /BF: Cp	RE: SEF3 /BF: SEF3	RE: RTBP1	RE: E RE: P2	RE: PB /BF: WPBF RE: P, hor2 / BF: BLZ1 , PBF RE: P /BF: PBF1 (2)	RE: EIN3 /BF: EIN3	RE: WUS /BF: WUS	RE: T/G /BF: GL3, DEL65	RE: GARE1	RE: AuxRE	RE: DOF1	RE: SEF3 /BF: SEF3 RE: SEF3 -2 /BF: SEF3
Nsite-PL	2_4	50	630	25-AGGGG-21 (2)	25-TAAAATAT-32		28-TTTAGGG-22						73-TAACGTAG-66			
		51	712					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)					81- ATGTCCCAT-89		91-AACCCA-96 (2)
	4_5	1	712					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)					81- ATGTCCCAT-89		91-AACCCA-96 (2)
	8_2	4	707					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)			45-AACGTG-50 (3)				
	18_1	1	636													78-AACCCA-83 (2)
	19_1	4	707			28-AACCCA-33 (2)				45-ATGTATAC-38	49-TTTAATGT-42				94-GAAAGG-89	95-AACCCA-100 (2)
	20_2	1	707			28-AACCCA-33 (2)				45-ATGTATAC-38	49-TTTAATGT-42				94-GAAAGG-89	95-AACCCA-100 (2)
	26_L5.S.R.	1	712					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)					81- ATGTCCCAT-89		91-AACCCA-96 (2)
	27_2	1	707					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)			45-AACGTG-50 (3)				
		47	700					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)			45-AACGTG-50 (3)				
		52	707					41-TGTAAAGT-34 (2)	41-TGTAAAG-35 (4)			45-AACGTG-50 (3)				
				1	2	3	4	5	6	7	8	9	10	11	12	13

	Cell line	Clone	Lenght (bp)	RE: G /BF: RITA1, bZIP	RE: ABRE	RE: Em1b /BF: EmBP1	RE: AuxRE	RE: Box III	RE: Amylase RE: TATCCAT/C RE: Box I /BF: aleurone	RE: E / G /BF: seed protein RE: NDE 1	RE: E / G /BF: seed protein RE: NDE1	RE: A1 /BF: CBF1	RE: A1 /BF: CBF1	RE: BN /BF: GT1	RE: W /BF: WRKY	RE: I-box
ScanWM-P	2_4	50	630											212-ATTTTGTAATTT-223	365-TCTGACT-359	385-GATAAGA-391
		51	712	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54	81-ATGTCCCAT-89	96-TGGGTTAATG-87	108-TATCCAT-114 (3)					304-ATTTTGTAATTT-315		
	4_5	1	712	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54	81-ATGTCCCAT-89	96-TGGGTTAATG-87	108-TATCCAT-114 (3)					304-ATTTTGTAATTT-315		
	8_2	4	707	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54			91-TATCCAT-97 (3)					299-ATTTTGTAATTT-310		
	18_1	1	636											225-ATTTTGTAATTT-236		
	19_1	4	707							115-CATATG-120 (2)	120-CATATG-115 (2)	169-TCCGAC-164	211-TACGAC- 206			
	20_2	1	707							115-CATATG-120 (2)	120-CATATG-115 (2)	169-TCCGAC-164	211-TACGAC- 206			
	26_L5.S.R.	1	712	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54	81-ATGTCCCAT-89	96-TGGGTTAATG-87	108-TATCCAT-114 (3)					304-ATTTTGTAATTT-315		
	27_2	1	707	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54			91-TATCCAT-97 (3)					299-ATTTTGTAATTT-310		
		47	700	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54			91-TATCCAT-97 (3)					292-ATTTTGTAATTT-303		
		52	707	45-AACGTG-50	46-ACGTG-50	46-ACGTGTATT-54			91-TATCCAT-97 (3)					299-ATTTTGTAATTT-310		
				1	2	3	4	5	6	7	8	9	10	11	12	13

	Cell line	Clone	Lenght (bp)	RE: CM2 /BF: CAMTA3	RE: PRE1 /BF: AGAMOUS	RE: GA5 /BF: BPC1	RE: CT-rich	RE: GA2 /BF: BPC1	RE: GA5 /BF: BPC1	RE: CT-rich	RE: GT1 K	RE: P2	RE: CCA1 /BF: CCA1	RE: TAGTCAAC	RE: E2/BF: DPBF1/2 RE: distal /BF: ESE1	RE: EE /BF: RVE1/2/3/4/7/8 (2), LHY, TLP11 RE: EE /BF: CCA1 (2)
Nsite-PL	2_4	47	2071	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
	4_5	8	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
	8_2	3	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
		17	2068	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288						736-CCACTTG-742 (2)	
	18_1	S1	2076	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (2-1)		202-AGAAAGAGA-210	279-AGAGAGAGA-287 (2-1)		311-AGGCATT-305				737-CCACTTG-743 (2)	
		L56	1981	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (6-2)	191-CTCTCTCTCTCTCTCTCT-176 (3-1)	212-AGAAAGAGA-220	289-AGAGAGAGA-297 (6-2)	304-CTCTCTCTCTCTCTCT-289 (2-1)						
		L146	1957	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (2-1)		202-AGAAAGAGA-210	279-AGAGAGAGA-287 (2-1)		311-AGGCATT-305				737-CCACTTG-743 (2)	
		L177	1979	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (5-1)	191-CTCTCTCTCTCTCTCT-176 (2-1)	210-AGAAAGAGA-218	287-AGAGAGAGA-295 (5-1)	302-CTCTCTCTCTCTCTCT-287 (2-1)						
		L214	1980	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (6-2)	191-CTCTCTCTCTCTCTCT-176 (3-1)	212-AGAAAGAGA-220	289-AGAGAGAGA-297 (4-2)	304-CTCTCTCTCTCTCTCT-289						
		L219	1957	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (2-1)		212-AGAAAGAGA-220	279-AGAGAGAGA-287 (2-1)		311-AGGCATT-305				737-CCACTTG-743 (2)	
	19_1	4	1987			192-AGAGAGAGA-200 (7-2)				207-CTCTCTCTCTCTCTCT-192 (2-1)		318-TGTAAAGT-325 (2)	448-AAGAATCTA-440	567-TAGTCAAC-574		883-AAAATATCT-875 (3)
		14	1989			192-AGAGAGAGA-200 (8-2)				207-CTCTCTCTCTCTCTCT-192 (5-1)		320-TGTAAAGT-327 (2)	450-AAGAATCTA-442	569-TAGTCAAC-576		885-AAAATATCT-877 (3)
		64	1985			192-AGAGAGAGA-200 (6-2)				207-CTCTCTCTCTCTCTCT-192 (3-1)		316-TGTAAAGT-323 (2)	446-AAGAATCTA-438	565-TAGTCAAC-572		881-AAAATATCT-873 (3)
	20_2	3	1987			192-AGAGAGAGA-200 (8-2)				207-CTCTCTCTCTCTCTCT-192 (5-1)		320-TGTAAAGT-327 (2)	450-AAGAATCTA-442	569-TAGTCAAC-576		885-AAAATATCT-877 (3)
		31	1984			192-AGAGAGAGA-200 (6-2)				207-CTCTCTCTCTCTCTCT-192 (3-1)		316-TGTAAAGT-323 (2)	446-AAGAATCTA-438	565-TAGTCAAC-572		881-AAAATATCT-873 (3)
		35	1985			192-AGAGAGAGA-200 (7-2)				207-CTCTCTCTCTCTCTCT-192 (4-1)		318-TGTAAAGT-325 (2)	448-AAGAATCTA-440	567-TAGTCAAC-574		883-AAAATATCT-875 (3)
	26_L5.S.R.	S5	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
		L13	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
		L101	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
		L128	2061	102-CCGCGT-107	131-AAAGGCAA-124	176-AGAGAGAGA-184 (2-1)		202-AGAAAGAGA-210	261-AGAGAGAGA-269 (2-1)						715-CCACTTG-721 (2)	
		L142	2079	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	278-AGAGAGAGA-286 (2-1)						732-CCACTTG-738 (2)	
	27_2	27	2070	101-CCGCGT-106	130-AAAGGCAA-123	175-AGAGAGAGA-183		201-AGAAAGAGA-209	280-AGAGAGAGA-288 (2-1)						738-CCACTTG-744 (2)	
				1	2	3	4	5	6	7	8	9	10	11	12	13

E — *AOX2a intron 1* — Nsite-PL / ScanWM-P

	Cell line	Clone	Lenght (bp)	RE: Box V	RE: ABRE B /BF: ABF	RE: BOX A2	RE: A1 /BF: CBF1	RE: Box L	RE: A1 /BF: CBF1	RE: A1 /BF: CBF1	RE: ATCATC	RE: TGA1	RE: A1 /BF: CBF1	RE: A1 /BF: CBF1	RE: CRE	RE: W /BF: WRKY
Nsite-PL	2_4	47	2071	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
	4_5	8	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
	8_2	3	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
		17	2068	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	324-ATCATC-329			387-ATCGAC-392	580-CAATCAGATC-589	
	18_1	S1	2076		101-TCCGCGTCTC-110			198-TCTCCCCTTCC-188		270-ATCGAC-265				388-ATCGAC-393	581-CAATCAGATC-590	
		L56	1981	53-GTTTTTTCTTGTCTGTT-69	101-TCCGCGTCTC-110			208-TCTCCCCTTCC-198		280-ATCGAC-275	341-ATCATC-346			404-ATCGAC-409	597-CAATCAGATC-606	
		L146	1957		101-TCCGCGTCTC-110			198-TCTCCCCTTCC-188		270-ATCGAC-265				388-ATCGAC-393	581-CAATCAGATC-590	
		L177	1979	53-GTTTTTTCTTGTCTGTT-69	101-TCCGCGTCTC-110			206-TCTCCCCTTCC-196		278-ATCGAC-273	339-ATCATC-344			402-ATCGAC-407	595-CAATCAGATC- 604	
		L214	1980	53-GTTTTTTCTTGTCTGTT-69	101-TCCGCGTCTC-110			208-TCTCCCCTTCC-198		280-ATCGAC-275	339-ATCATC-344			402-ATCGAC-407	595-CAATCAGATC-604	
		L219	1957		101-TCCGCGTCTC-110			198-TCTCCCCTTCC-188		270-ATCGAC-265				388-ATCGAC-393	581-CAATCAGATC-590	
	19_1	4	1987			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		240-CTCGAC-235			298-TGACGGAC-291	299-CTCGAC-304			505-TCTGACT-499
		14	1989			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		242-CTCGAC-237			300-TGACGGAC-293	301-CTCGAC-306			507-TCTGACT-501
		64	1985			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		238-CTCGAC-233			296-TGACGGAC- 289	297-CTCGAC-302			503-TCTGACT- 497
	20_2	3	1987			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		242-CTCGAC-237			300-TGACGGAC-293	301-CTCGAC-306			507-TCTGACT-501
		31	1984			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		238-CTCGAC-233			296-TGACGGAC- 289	297-CTCGAC-302			503-TCTGACT- 497
		35	1985			113-TCGATTTCAATTTCCTTTT-131	173-CGCGAC-178		240-CTCGAC-235			298-TGACGGAC-291	299-CTCGAC-304			505-TCTGACT-499
	26_L5.S.R.	S5	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
		L13	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
		L101	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
		L128	2061		101-TCCGCGTCTC-110			198-TCTCCCCTTCC-188		252-ATCGAC- 247	307-ATCATC-312			370-ATCGAC-375	563-CAATCAGATC- 572	
		L142	2079	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	324-ATCATC-329			387-ATCGAC-392	580-CAATCAGATC-589	
	27_2	27	2070	53-GTTTTTTCTTGTCTGTT-69	100-TCCGCGTCTC-109			197-TCTCCCCTTCC-187		269-ATCGAC-264	326-ATCATC-331			389-ATCGAC-394	582-CAATCAGATC-591	
				1	2	3	4	5	6	7	8	9	10	11	12	13

F — *AOX2b intron 1* — Nsite-PL / ScanWM-P

	Cell line	Clone	Length (bp)	RE: CCA1 /BF: CCA1	RE: AC-1 /BF: MYB61	RE: ABRE /BF: ABI3	RE: CE1 /BF: ABI4	RE: CRE 45-48	RE: ZPT2 /BF: ZPT2-2	RE: WI /BF: MINI3	RE: GAAATA	RE: Inr	RE: DRE1	RE: ICE2	RE: NF-kB	RE: CRE 33-36/39-42
Nsite-PL	2_4	57	1019		18-ACCACC-13	23-CAAACACC-16	24-GCCACCG-30	31-GCGGTGG- 25	78-CAGT- 75 --10-- 64-TGCT-61	95-TTGACTT- 89	154-GGAAATAA-147	201-TTTCACTT-208				
	4_5	1_16	823	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
		3_21	1019		18-ACCACC-13	23-CAAACACC-16	24-GCCACCG-30	31-GCGGTGG- 25	78-CAGT- 75 --10-- 64-TGCT-61	95-TTGACTT- 89	154-GGAAATAA-147	201-TTTCACTT-208				
	8_2	0_32	560	7-AACAATCT-14									51-ACCGAGA-57		92-GGACTTTTC-100	
		1_34	823	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
		1_45	823	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
	18_1	17	560	7-AACAATCT-14									51-ACCGAGA-57		92-GGACTTTTC-100	
	19_1	2_26	823											124-ACTCCG-129		130-TCGGAG-125
		2_31	823	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
	20_2	5	832	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
	26_L5.S.R.	17	822											124-ACTCCG-129		130-TCGGAG-125
	27_2	47	560	7-AACAATCT-14									51-ACCGAGA-57		92-GGACTTTTC-100	
		56	823	7-AACAATCT-14									84-ACCGAGA-90	122-ACTCCG-127		128-TCGGAG-123
		57	823											124-ACTCCG-129		130-TCGGAG-125
				1	2	3	4	5	6	7	8	9	10	11	12	13

	Cell line	Clone	Length (bp)	RE: ATCATC	RE: W /BF: WRKY	RE: NF-kB	RE: W /BF: WRKY	RE: ATCATC	RE: A1 /BF: CBF1 RE: R /BF: CBF1	RE: O2d /BF: O2	RE: O2d /BF: O2	RE: Box III /BF: GT1	RE: I-box	RE: AuxRE	RE: O2d /BF: O2
ScanWM-P	2_4	57	1019		96-CTTGACT-90		135-CATGACT-141	319-ATCATC-314	372-TCCGAC-377		420-CACGAC-425		488-GATAAGA-494		
	4_5	1_16	823	71-ATCATC-66				71-ATCATC-66	124-TCCGAC-129		172-CACGAC-177 (2)		241-GATAAGA-247		
		3_21	1019		96-CTTGACT-90		135-CATGACT-141	319-ATCATC-314	372-TCCGAC-377		420-CACGAC-425 (2)		488-GATAAGA-494		
	8_2	0_32	560			92-GGACTTTTC-100					138-CACGAC-143 (2)	202-AAGATAAGA-210	204-GATAAGA-210	329-ATGTCCCGT-337	333-GACATGGCAT-324
		1_34	823	71-ATCATC-66					124-TCCGAC-129		172-CACGAC-177 (2)		241-GATAAGA-247		
		1_45	823	71-ATCATC-66					124-TCCGAC-129		172-CACGAC-177 (2)		241-GATAAGA-247		
	18_1	17	560			92-GGACTTTTC-100					138-CACGAC-143 (2)	202-AAGATAAGA-210	204-GATAAGA-210	329-ATGTCCCGT-337	333-GACATGGCAT-324
	19_1	2_26	823	73-ATCATC- 68					126-TCCGAC-131						
		2_31	823	71-ATCATC-66					124-TCCGAC-129		172-CACGAC-177 (2)		241-GATAAGA-247		
	20_2	5	832	71-ATCATC-66					124-TCCGAC-129	170-GACATGACAT-179		177-GTCATGTCAT-168	240-GATAAGA-246		
	26_L5.S.R.	17	822	73-ATCATC- 68					126-TCCGAC-131						
	27_2	47	560			92-GGACTTTTC-100					138-CACGAC-143 (2)	202-AAGATAAGA-210	204-GATAAGA-210	329-ATGTCCCGT-337	333-GACATGGCAT-324
		56	823	71-ATCATC-66					124-TCCGAC-129		172-CACGAC-177 (2)		241-GATAAGA-247		
		57	823	73-ATCATC- 68					126-TCCGAC-131						
				1	2	3	4	5	6	7	8	9	10	11	12	13

D — *AOX1 intron 1*

	Cell line	Clone	Lenght (bp)	RE: CCAAAT /BF: ROM1/2	RE: Amylase RE: Box I /BF: aleurone	RE: WUS /BF: WUS RE: WOX11 /BF: WOX11	RE: GT1 /BF: GT2	RE: C2a /BF: root factor	RE: FLS2 EIN3/EIL1 /BF: EIN3/EIL1	RE: 16 KK RE: CCGTTA	RE: REbeta	RE: Box I /BF: aleurone RE: Amylase	RE: PacC BS /BF: PacC RE: GCCAAG	RE: UV /BF: NSP1	RE: REbeta	RE: P1 /BF: PHR1
Nsite-PL	2_4	50	630					94-GCTTATC-100					116-GCCAAG-121 (2)			
		51	712			92-TTAATGG-86 (2)	94-GGTTAAT-88					108-TATCCAT-114 (2)				159-GGATATTC-166
	4_5	1	712			92-TTAATGG-86 (2)	94-GGTTAAT-88					108-TATCCAT-114 (2)				159-GGATATTC-166
	8_2	4	707		91-TATCCAT-97 (2)									121-CCAAGG-126	141-CGGATA-146	142-GGATATTC-149
	18_1	1	636	81-CCAAAT-86						102-CCGTTA-107 (2)	104-CGGATA-99					
	19_1	4	707	98-CCAAAT-103					102-ATGTATGT-109						166-CGGATA-171	
	20_2	1	707	98-CCAAAT-103					102-ATGTATGT-109						166-CGGATA-171	
	26_L5.S.R.	1	712			92-TTAATGG-86 (2)	94-GGTTAAT-88					108-TATCCAT-114 (2)				159-GGATATTC-166
	27_2	1	707		91-TATCCAT-97 (2)									121-CCAAGG-126	141-CGGATA-146	142-GGATATTC-149
		47	700		91-TATCCAT-97 (2)									121-CCAAGG-126		
		52	707		91-TATCCAT-97 (2)									121-CCAAGG-126	141-CGGATA-146	142-GGATATTC-149
				14	15	16	17	18	19	20	21	22	23	24	25	26

	Cell line	Clone	Lenght (bp)	RE: PRD2 /BF: nuclear protein	RE: W /BF: WRKY	RE: Box II /BF: TCP20 (2)	RE: W /BF: WRKY	RE: W /BF: WRKY RE: EIRE /BF: WRKY1/2/3	RE: NDE 1 RE: E /G /BF: seed protein	RE: NDE Box 1 RE: E /G /BF: seed protein	RE: ocs /BF: OCSTF	RE: Ce3	RE: W /BF: WRKY	RE: W /BF: WRKY	RE: EIRE /BF: WRKY1/2/3
ScanWM-P	2_4	50	630	426-ATCCATATTTTACTC-440								483-ACAAATGTCCTC-472			
		51	712		449-CCTGACT-443		506-GTTGACC-512	507-TTGACC-512				567-ACAAATGTCCTC-556	618-AGTGACC-624		
	4_5	1	712		449-CCTGACT-443		506-GTTGACC-512	507-TTGACC-512				567-ACAAATGTCCTC-556	618-AGTGACC-624		
	8_2	4	707		444-CCTGACT-438		501-GTTGACC-507	502-TTGACC-507				562-ACAAATGTCCTC-551	613-AGTGACC-619		
	18_1	1	636		371-CCTGACT-365		430-GTTGACC-436	431-TTGACC-436	485-CATATG-490 (2)	490-CATATG-485 (2)				542-ATTGACC-548	543-TTGACC-548
	19_1	4	707		439-CCTGACT-433	504-TGGGCC-499 (2)					548-TGACACTTGTTTTTACGTTT-567	557- ACAAGTGTCATC- 546		613-ATTGACT-619	
	20_2	1	707		439-CCTGACT-433	504-TGGGCC-499 (2)					548-TGACACTTGTTTTTACGTTT-567	557- ACAAGTGTCATC- 546		613-ATTGACT-619	
	26_L5.S.R.	1	712		449-CCTGACT-443		506-GTTGACC-512	507-TTGACC-512				567-ACAAATGTCCTC-556	618-AGTGACC-624		
	27_2	1	707		444-CCTGACT-438		501-GTTGACC-507	502-TTGACC-507				562-ACAAATGTCCTC-551		613-ATTGACT-619	
		47	700		437-CCTGACT-431			495-TTGACC-500 (2)				555-ACAAATGTCCTC-544	606-AGTGACC-612		
		52	707		444-CCTGACT-438		501-GTTGACC-507	502- TTGACC-507				562-ACAAATGTCCTC-551		613-ATTGACT-619	
				14	15	16	17	18	19	20	21	22	23	24	25

E — *AOX2a intron 1*

	Cell line	Clone	Lenght (bp)	RE: KN1/KIP /BF: KN1/KIP	RE:AC 1 /BF: MYB1/4	RE: GT2 /BF: DF1	RE: AGCC /BF: ERF1	RE: RIN /BF: RIN	RE: EM1 /BF: MADS	RE: G /BF: GBP	RE: ABRE 1/2 /BF: ABI3/5, AREB1	RE: H21 /BF: H21	RE: HSE4 /BF: HsfA2	RE: Inr	RE: LFY /BF: LFY	RE: RY
Nsite-PL	2_4	47	2071	1150-TGACAGGT-1157	1162-ACCTACA-1168					1466-CCACGTCG-1473	1471-ACGTGGC-1465	1476-TGTGACAGT-1484	1540-GAATAATTC-1548	1562-TTTCACTT-1569		1669-CATGCAC-1675
	4_5	8	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
	8_2	3	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
		17	2068	1148-TGACAGGT-1155	1160-ACCTACA-1166					1463-CCACGTCG-1470	1468-ACGTGGC-1462	1473-TGTGACAGT-1481	1537- GAATAATTC-1545	1559- TTTCACTT-1566		1666-CATGCAC-1672
	18_1	S1	2076	1149-TGACAGGT-1156	1161-ACCTACA-1167		1207-GCAGCC-1212			1471-CCACGTCG-1478	1476-ACGTGGC-1470	1481-TGTGACAGT-1489	1545-GAATAATTC-1553	1567-TTTCACTT-1574		1674-CATGCAC-1680
		L56	1981	1173-TGACAGGT-1180	1185-ACCTACA-1191					1376-CCACGTCG-1383	1381-ACGTGGC-1375	1386-TGTGACAGT-1394	1450-GAATAATTC-1458	1472-TTTCACTT-1479		1579-CATGCAC-1585
		L146	1957	1149-TGACAGGT-1156	1161-ACCTACA-1167				1295-CTTTTTTTGG-1304 (2)	1352-CCACGTCG-1359	1357-ACGTGGC-1351	1362-TGTGACAGT-1370	1426-GAATAATTC-1434	1448-TTTCACTT-1455		1555-CATGCAC-1561
		L177	1979	1171-TGACAGGT-1178	1183-ACCTACA-1189					1374-CCACGTCG-1381	1379-ACGTGGC-1373	1384-TGTGACAGT-1392	1448-GAATAATTC-1456	1470-TTTCACTT-1477		1577-CATGCAC-1583
		L214	1980	1171-TGACAGGT-1178	1183-ACCTACA-1189					1375-CCACGTCG-1382	1380-ACGTGGC-1374	1385-TGTGACAGT-1393	1449-GAATAATTC1457	1471-TTTCACTT-1478		1578-CATGCAC-1584
		L219	1957	1149-TGACAGGT-1156	1161-ACCTACA-1167					1352-CCACGTCG-1359	1357-ACGTGGC-1351	1362-TGTGACAGT-1370	1426-GAATAATTC-1434	1448-TTTCACTT-1455		1555-CATGCAC-1561
	19_1	4	1987	1045-TGACAGCT-1052				1231-CTAGATTTAG-1240							1554-CCAATGT-1560	
		14	1989	1047-TGACAGCT-1054		1131-GGTAATTA-1138		1233-CTAGATTTAG-1242							1556-CCAATGT-1562	
		64	1985	1043-TGACAGCT-1050		1127-GGTAATTA-1134		1229-CTAGATTTAG-1238							1552-CCAATGT-1558	
	20_2	3	1987	1047-TGACAGCT-1054		1131-GGTAATTA-1138		1231-CTAGATTTAG-1240							1554-CCAATGT-1560	
		31	1984	1043-TGACAGCT-1050		1127-GGTAATTA-1134		1228-CTAGATTTAG-1237							1551-CCAATGT-1557	
		35	1985	1045-TGACAGCT-1052		1129-GGTAATTA-1136		1229-CTAGATTTAG-1238							1552-CCAATGT-1558	
	26_L5.S.R.	S5	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
		L13	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
		L101	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
		L128	2061	1127-TGACAGGT-1134			1185-GCAGCC-1190			1456-CCACGTCG-1463	1461-ACGTGGC-1455		1530-GAATAATTC-1538	1552-TTTCACTT-1559		
		L142	2079	1144-TGACAGGT-1151			1202-GCAGCC-1207			1474-CCACGTCG-1481	1479-ACGTGGC-1473		1548-GAATAATTC-1556	1570-TTTCACTT-1577		
	27_2	27	2070	1150-TGACAGGT-1157	1162-ACCTACA-1168					1465-CCACGTCG-1472	1470-ACGTGGC-1464	1475-TGTGACAGT-1483	1539-GAATAATTC-1547	1561-TTTCACTT-1568		1668-CATGCAC-1674
				14	15	16	17	18	19	20	21	22	23	24	25	26

	Cell line	Clone	Lenght (bp)	RE: ATCATC	RE: ATCATC	RE: W /BF: WRKY	RE: Element 1 /BF: nodule factor	RE: ERE /BF: WRKY1/2/4	RE: W /BF: WRKY	RE: W /BF: WRKY	RE: E2 /BF: DPBF1/2	RE: W /BF: WRKY	RE: Box III	RE:A1 /BF: CBF1	RE: ERE /BF: WRKY1/2/4	RE: AT1
ScanWM-P	2_4	47	2071		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
	4_5	8	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
	8_2	3	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
		17	2068		603-ATCATC-598			647-TTTAGTCAGCTAATCAC- 663	656-GCTGACT-650		736-CCACTTG-742	754-AATGACT-760				
	18_1	S1	2076		604-ATCATC-599			648-TTTAGTCAGCTAATCAC- 664	657-GCTGACT- 651		737-CCACTTG-743			755-AACGAC-760		
		L56	1981		620-ATCATC-615			664-TTTAGTCAGCTAATCAC-680	673-GCTGACT- 667			762-AATGACT-768		779-AACGAC-784		
		L146	1957		604-ATCATC-599			648-TTTAGTCAGCTAATCAC- 664	657-GCTGACT- 651		737-CCACTTG- 743			755-AACGAC-760		
		L177	1979		618-ATCATC- 613			662-TTTAGTCAGCTAATCAC- 678	671-GCTGACT- 665			760-AATGACT-766	762-TGACTTAATG-771	777-AACGAC-782		
		L214	1980		618-ATCATC- 613			662-TTTAGTCAGCTAATCAC- 678	671-GCTGACT- 665			760-AATGACT-766	762-TGACTTAATG-771	777-AACGAC-782		
		L219	1957		604-ATCATC-599			648-TTTAGTCAGCTAATCAC- 664	657-GCTGACT- 651		737-CCACTTG- 743			755-AACGAC-760		
	19_1	4	1987	517-ATCATC-522		574-GTTGACT- 568	598-TATATATTAATACATAATTAGATT-621			744-CATGACT-738					749-GAACCCATG-741	793-TAATAATTTCATA-781
		14	1989	519-ATCATC-524		576-GTTGACT-570	600-TATATATTAATACATAATTAGATT-623			746-CATGACT-740					751-GAACCCATG-743	795-TAATAATTTCATA-783
		64	1985	515-ATCATC-520		572-GTTGACT-566	596-TATATATTAATACATAATTAGATT-619			742-CATGACT-736					747-GAACCCATG-739	791-TAATAATTTCATA-779
	20_2	3	1987	519-ATCATC-524		576-GTTGACT-570	600-TATATATTAATACATAATTAGATT-623			746-CATGACT-740					751-GAACCCATG-743	795-TAATAATTTCATA-783
		31	1984	515-ATCATC-520		572-GTTGACT-566	596-TATATATTAATACATAATTAGATT-619			742-CATGACT-736					747-GAACCCATG-739	791-TAATAATTTCATA-779
		35	1985	517-ATCATC-522		574-GTTGACT- 568	598-TATATATTAATACATAATTAGATT-621			744-CATGACT-738					749-GAACCCATG-741	793-TAATAATTTCATA-781
	26_L5.S.R.	S5	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
		L13	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
		L101	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
		L128	2061		586-ATCATC-581			626-TTTAGTCAGCTAATCAC-642	635-GCTGACT- 629		715-CCACTTG-721			733-AACGAC-738		
		L142	2079		603-ATCATC-598			643-TTTAGTCAGCTAATCAC-659	652-GCTGACT- 646		732-CCACTTG-738			750-AACGAC-755		
	27_2	27	2070		605-ATCATC-600			649-TTTAGTCAGCTAATCAC-665	658-GCTGACT-652		738-CCACTTG-744	756-AATGACT-762				
				14	15	16	17	18	19	20	21	22	23	24	25	26

F — *AOX2b intron 1*

	Cell line	Clone	Length (bp)	RE: H21 /BF: H21	RE: H21 /BF: H21	RE: R	RE: TGTCACA	RE: TCGTGT	RE: PB /BF: WPBF	RE: CE1 /BF: ABI4	RE: 5A W	RE: Box III /BF: GT1	RE: I-box	RE: - 300	RE: SMRE1 /BF: MYB3/2/20/21	RE: AC-I /BF: MYB61
Nsite-PL	2_4	57	1019	399-TCTGACACT-407	415-TGTGACACG-423	420-CACGAC- 425	421-TGTCACA- 415						488-GATAAGA- 494	503-TGTAGAA-509	516-ACCAAAT-522	
	4_5	1 16	823	151-TCTGACACT-159	167-TGTGACACG-175	172-CACGAC-177	173-TGTCACA-167			210-CACCGA- 205			241-GATAAGA-247	256-TGTAGAA- 262	269-ACCAAAT-275	
		3 21	1019	399-TCTGACACT-407	415-TGTGACACG-423	420-CACGAC- 425	421-TGTCACA- 415						488-GATAAGA- 494	503-TGTAGAA-509	516-ACCAAAT-522	
	8_2	0 32	560	117-TCTGACACT-125	133-TGTGACACG-141	138-CACGAC-143	139-TGTCACA-133	142-TCGTGT-137	162-TGTAAAG-168 (4)		190-CCAAAAA-196	202-AAGATAAGA- 210	204-GATAAGA-210			300-ACCACA-305
		1 34	823	151-TCTGACACT-159	167-TGTGACACG-175	172-CACGAC-177	173-TGTCACA-167			210-CACCGA- 205			241-GATAAGA-247	256-TGTAGAA- 262	269-ACCAAAT-275	
		1 45	823	151-TCTGACACT-159	167-TGTGACACG-175	172-CACGAC-177	173-TGTCACA-167			210-CACCGA- 205			241-GATAAGA-247	256-TGTAGAA- 262	269-ACCAAAT-275	
	18_1	17	560	117-TCTGACACT-125	133-TGTGACACG-141	138-CACGAC-143	139-TGTCACA-133	142-TCGTGT-137	162-TGTAAAG-168 (4)		190-CCAAAAA-196	202-AAGATAAGA- 210	204-GATAAGA-210			300-ACCACA-305
	19_1	2 26	823	153-TCTGACAGT-161			175-TGTCACA-169			212-CACCGA-207					259-ACCAAAT-265	
		2 31	823	151-TCTGACACT-159	167-TGTGACACG-175	172-CACGAC-177	173-TGTCACA-167			210-CACCGA- 205			241-GATAAGA-247	256-TGTAGAA- 262	269-ACCAAAT-275	
	20_2	5	832	151-TCTGACACT-159						210-CACCGA- 205			240-GATAAGA-246	255-TGTAGAA-261	268-ACCAAAT-274	
	26_L5.S.R.	17	822	153-TCTGACACT-161			175-TGTCACA-169			212-CACCGA-207					258-ACCAAAT-264	
	27_2	47	560	117-TCTGACACT-125	133-TGTGACACG-141	138-CACGAC-143	139-TGTCACA-133	142-TCGTGT-137	162-TGTAAAG-168 (4)		190-CCAAAAA-196	202-AAGATAAGA- 210	204-GATAAGA-210			300-ACCACA-305
		56	823	151-TCTGACACT-159	167-TGTGACACG-175	172-CACGAC-177	173-TGTCACA-167			210-CACCGA- 205			241-GATAAGA-247	256-TGTAGAA- 262	269-ACCAAAT-275	
		57	823	153-TCTGACACT-161			175-TGTCACA-169			212-CACCGA-207					259-ACCAAAT-265	
				14	15	16	17	18	19	20	21	22	23	24	25	26

	Cell line	Clone	Length (bp)	RE: Box A	RE: B1 /BF: SGBF1/2	RE: ATF /BF: AP1	RE: ABRE2	RE: bZIP	RE: ABRE	RE: Em1b /BF: EmBP1	RE: Em1b /BF: EmBP1	RE: G /BF: RITA1, bZIP RE: ABRE1	RE: TGA1	RE: SA/MJ-RE	RE: W /BF: WRKY	RE: ATCATC
ScanWM-P	2_4	57	1019			644-ACACGTCA- 651	645-CACGTC-650	646-ACGTCA-651	649-ACGTG-645		649-ACGTGTAAG-641	650-GACGTG- 645 (3)	651-TGACGTGT- 644	651-TGACGT-646		691-ATCATC- 686
	4_5	1 16	823			458-ACACGTCA-465	459-CACGTC-464	460-ACGTCA-465	463-ACGTG-459	463-ACGTGTTGG-455		464-GACGTG-459 (2)	465-TGACGTGT-458	465-TGACGT-460	490-GATGACT-496	
		3 21	1019			644-ACACGTCA- 651	645-CACGTC-650	646-ACGTCA-651	649-ACGTG-645		649-ACGTGTAAG-641	650-GACGTG- 645 (2)	651-TGACGTGT- 644	651-TGACGT-646		691-ATCATC- 686
	8_2	0 32	560	334-CCGTCC-339	337-TCCCCGTGTC-346											
		1 34	823			458-ACACGTCA-465	459-CACGTC-464	460-ACGTCA-465	463-ACGTG-459	463-ACGTGTTGG-455		464-GACGTG-459 (2)	465-TGACGTGT-458	465-TGACGT-460	490-GATGACT-496	
		1 45	823			458-ACACGTCA-465	459-CACGTC-464	460-ACGTCA-465	463-ACGTG-459	463-ACGTGTTGG-455		464-GACGTG-459 (2)	465-TGACGTGT-458	465-TGACGT-460	490-GATGACT-496	
	18_1	17	560	334-CCGTCC-339	337-TCCCCGTGTC-346											
	19_1	2 26	823			448-ACACGTCA-455	449-CACGTC-454	450-ACGTCA- 455	453-ACGTG-449	453-ACGTGTTGG-445		454-GACGTG-449 (2)	455-TGACGTGT- 448	455-TGACGT-450		495-ATCATC-490
		2 31	823			458-ACACGTCA-465	459-CACGTC-464	460-ACGTCA-465	463-ACGTG-459	463-ACGTGTTGG-455		464-GACGTG-459 (2)	465-TGACGTGT-458	465-TGACGT-460	490-GATGACT-496	
	20_2	5	832			457-ACACGTCA-464	458-CACGTC-463	459-ACGTCA-464	462-ACGTG-458	462-ACGTGTTGG-454		463-GACGTG-458 (2)	464-TGACGTGT-457	464-TGACGT-459		504-ATCATC-499
	26_L5.S.R.	17	822			447-ACACGTCA-454	448-CACGTC-453	449-ACGTCA-454	452-ACGTG-448	452-ACGTGTTGG-444		453-GACGTG-448 (2)	454-TGACGTGT-447	454-TGACGT-449		494-ATCATC-489
	27_2	47	560	334-CCGTCC-339	337-TCCCCGTGTC-346											
		56	823			458-ACACGTCA-465	459-CACGTC-464	460-ACGTCA-465	463-ACGTG-459	463-ACGTGTTGG-455		464-GACGTG-459 (2)	465-TGACGTGT-458	465-TGACGT-460	490-GATGACT-496	
		57	823			448-ACACGTCA-455	449-CACGTC-454	450-ACGTCA- 455	453-ACGTG-449	453-ACGTGTTGG-445		454-GACGTG-449 (2)	455-TGACGTGT- 448	455-TGACGT-450		495-ATCATC-490
				14	15	16	17	18	19	20	21	22	23	24	25	26

D

AOX1 intron 1

Nste-PL

Cell line	Clone	Lenght (bp)	RE: P1 /BF: PHR1	RE: CCAAAT/BF: ROM1/2	RE: Cp /BF: Cp	RE: GTAC2 /BF: CRR1 RE: GTAC3 /BF: CRR1, SBP (2)	RE: GT1 K	RE: GAMYB /BF: GAMYB	RE: CAT	RE: GTAC2 /BF: CRR1 RE: GTAC3 /BF: CRR1, SBP (2)	RE: CARE D	RE: E5 /BF: DPBF1/2	RE: DOF	RE: W3 /BF: MINI3
2_4	**50**	630			131-TAAAATAT-138	161-TGTACC-156 (3)	168-AGGCATT-174			209-TGTACC-204 (3)				
	51	712	166-GAATATCC-159								212-CACACA-217	215-ACATCTG-221		
4_5	**1**	712	166-GAATATCC-159								212-CACACA-217	215-ACATCTG-221		
8_2	**4**	707	149-GAATATCC-142		171-TAAAATAT-178			184-TAACCAAT-177			207-CACACA-212	210-ACATCTG-216		
18_1	**1**	636			146-TAAAATAT-153				207-GCCAAC-202					
19_1	**4**	707		190-CCAAAT-195	196-TAAAATAT-203								235-GAAAGG-230	265-TTGACAA-259
20_2	**1**	707		190-CCAAAT-195	196-TAAAATAT-203								235-GAAAGG-230	265-TTGACAA-259
26_L5.S.R.	**1**	712	166-GAATATCC-159								212-CACACA-217	215-ACATCTG-221		
27_2	**1**	707	149-GAATATCC-142		171-TAAAATAT-178			184-TAACCAAT-177			207-CACACA-212	210-ACATCTG-216		
	47	700			164-TAAAATAT-171			177-TAACCAAT-170			200-CACACA-205	203-ACATCTG-209		
	52	707	149-GAATATCC-142		171-TAAAATAT-178			184-TAACCAAT-177			207-CACACA-212	210-ACATCTG-216		
			27	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**

ScanWM-P

Cell line	Clone	Lenght (bp)
2_4	**50**	630
	51	712
4_5	**1**	712
8_2	**4**	707
18_1	**1**	636
19_1	**4**	707
20_2	**1**	707
26_L5.S.R.	**1**	712
27_2	**1**	707
	47	700
	52	707

Nste-PL

Cell line	Clone	Lenght (bp)	RE: Inr	RE: Fp12/III	RE: H21 /BF: H21	RE: CM4 /BF: CAMTA3	RE: G /BF: MYC2	RE: MYC2/3/4 /BF: MYC2/3/4	RE: FHY3/FAR1 /BF: FHY3/FAR1 (2)	RE: CAMTA1/5 /BF: CAMTA1/5
2_4	**47**	2071	1681-TTTCAAAG-1688	1801-ATTTTTTTAT-1792				1949-CACGCG-1954	1949-CACGCGC-1955 (2)	1950-ACGCGC-1955
4_5	**8**	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
8_2	**3**	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
	17	2068	1678-TTTCAAAG-1685	1798-ATTTTTTTAT-1789				1946-CACGCG-1951	1946-CACGCGC-1952 (2)	1947-ACGCGC-1952
18_1	**S1**	2076	1686-TTTCAAAG-1693	1806-ATTTTTTTAT-1797				1954-CACGCG-1959	1954-CACGCGC-1960 (2)	1955-ACGCGC-1960
	L56	1981	1591-TTTCAAAG-1598	1711-ATTTTTTTAT-1702				1859-CACGCG-1864	1859-CACGCGC-1865 (2)	1860-ACGCGC-1865
	L146	1957	1567-TTTCAAAG-1574	1687-ATTTTTTTAT-1678				1835-CACGCG-1840	1835-CACGCGC-1841 (2)	1836-ACGCGC-1841
	L177	1979	1589-TTTCAAAG-1596	1709-ATTTTTTTAT-1700				1857-CACGCG-1862	1857-CACGCGC-1863 (2)	1858-ACGCGC-1863
	L214	1980	1590-TTTCAAAG-1597	1710-ATTTTTTTAT-1701				1858-CACGCG-1863	1858-CACGCGC-1864 (2)	1859-ACGCGC-1864
	L219	1957	1567-TTTCAAAG-1574	1687-ATTTTTTTAT-1678				1835-CACGCG-1840	1835-CACGCGC-1841 (2)	1836-ACGCGC-1841
19_1	**4**	1987			1786-TCTGACACT-1778	1864-TCCACGT-1870	1872-GCACGTGG-1865			
	14	1989			1788-TCTGACACT-1780	1866-TCCACGT-1872	1874-GCACGTGG-1867			
	64	1985			1784-TCTGACACT-1776	1862-TCCACGT-1868	1870-GCACGTGG-1863			
20_2	**3**	1987			1786-TCTGACACT-1778	1864-TCCACGT-1870	1872-GCACGTGG-1865			
	31	1984			1783-TCTGACACT-1775	1861-TCCACGT-1867	1869-GCACGTGG-1862			
	35	1985			1784-TCTGACACT-1776	1862-TCCACGT-1868	1870-GCACGTGG-1863			
26_L5.S.R.	**S5**	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
	L13	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
	L101	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
	L128	2061	1671-TTTCAAAG-1678	1791-ATTTTTTTAT-1782				1939-CACGCG-1944	1939-CACGCGC-1945 (2)	1940-ACGCGC-1945
	L142	2079	1689-TTTCAAAG-1696	1809-ATTTTTTTAT-1800				1957-CACGCG-1962	1957-CACGCGC-1963 (2)	1958-ACGCGC-1963
27_2	**27**	2070	1680-TTTCAAAG-1687	1800-ATTTTTTTAT-1791				1948-CACGCG-1953	1948-CACGCGC-1954 (2)	1949-ACGCGC-1954
			27	**28**	**29**	**30**	**31**	**32**	**33**	**34**

E

AOX2a intron 1

ScanWM-P

Cell line	Clone	Lenght (bp)	RE: GARE /Box 2 /BF: aleurone	RE: W /BF: WRKY	RE: W /BF: WRKY	RE: ABRE3a RE: G /BF: RITA1, bZIP	RE: ABRE	RE: Em1b /BF: EmBP1	RE: ABRE4	RE: Box II /BF: TCP20	RE: CRE	RE: ABRE	RE: Em1b /BF: EmBP1	RE: G /BF: RITA1, bZIP
2_4	**47**	2071	768-GAACAGAGTC-759	849-CGTGACT-843										
4_5	**8**	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
8_2	**3**	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
	17	2068	766-GAACAGAGTC-757	847-CGTGACT- 841										
18_1	**S1**	2076	767-GAACAGAGTC-758	848-CGTGACT-840										
	L56	1981	791-GAACAGAGTC-782	872-CGTGACT-866						1076-TGGGCT-1071				
	L146	1957	767-GAACAGAGTC-758	848-CGTGACT-840						1052-TGGGCT-1047				
	L177	1979	789-GAACAGAGTC- 780	870-CGTGACT- 864						1074-TGGGCT-1069				
	L214	1980	789-GAACAGAGTC- 780	870-CGTGACT- 864						1074-TGGGCT-1069				
	L219	1957	767-GAACAGAGTC-758	848-CGTGACT-840						1052-TGGGCT-1047				
19_1	**4**	1987			914-ACTGACT- 908	1041-TACGTG-1046 (2)	1042-ACGTG-1046	1042-ACGTGACAG-1050	1046-CACGTA-1041		1118-CAAACGAATC-1109			
	14	1989			916-ACTGACT- 910	1043-TACGTG-1048 (2)	1044-ACGTG-1048	1044-ACGTGACAG-1052	1048-CACGTA-1043		1120-CAAACGAATC-1111			
	64	1985			912-ACTGACT- 906	1039-TACGTG-1044 (2)	1040-ACGTG-1044	1040-ACGTGACAG-1048	1044-CACGTA-1039		1116-CAAACGAATC-1107			
20_2	**3**	1987			916-ACTGACT- 910	1043-TACGTG-1048 (2)	1044-ACGTG-1048	1044-ACGTGACAG-1052	1048-CACGTA-1043		1120-CAAACGAATC-1111			
	31	1984			912-ACTGACT- 906	1039-TACGTG-1044 (2)	1040-ACGTG-1044	1040-ACGTGACAG-1048	1044-CACGTA-1039		1116-CAAACGAATC-1107			
	35	1985			914-ACTGACT- 908	1041-TACGTG-1046 (2)	1042-ACGTG-1046	1042-ACGTGACAG-1050	1046-CACGTA-1041		1118-CAAACGAATC-1109			
26_L5.S.R.	**S5**	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
	L13	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
	L101	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
	L128	2061	745-GAACAGAGTC-736	826-CGTGACT-820								1153-ACGTG-1149	1153-ACGTGCCTT-1145	1154-AACGTG-1149
	L142	2079	762-GAACAGAGTC-753	843-CGTGACT-837								1170-ACGTG-1166	1170-ACGTGCCTT-1162	1171-AACGTG-1166
27_2	**27**	2070	768-GAACAGAGTC-759	849-CGTGACT-843										
			27	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**

F

AOX2b intron 1

Nste-PL

Cell line	Clone	Length (bp)	RE: Fp6/II	RE: Box A RE: CCGTCC (2)	RE: Fp6/II	RE: ERE 2 /BF: DOF	RE: ABRE2 (2) RE: PAN /BF: PAN, bZIP21, TGA6/5	RE: ABRE1 (2)	RE: GBF1 /BF: GBF1	RE: WUS 2 /BF: WUS	RE: ZPT2-2 /BF: ZPT2-2	RE: ERE2 /BF: DOF	RE: GTAC2 /BF: IPA1 (2)	RE: DREmut1 /BF: DREB3
2_4	**57**	1019	607-ACTCATA- 601						651-TGACGTGT- 644	662-AGTAATAT- 669	696-AGC- 694 -- 7-- 686-CAGT- 683	725-AATTCAAG-718		
4_5	**1_16**	823	360-ACTCATA-354		391-ACTCATA-385	405-AATTCAAG-398	459-CACGTC- 464 (3)	464-GACGTG- 459 (2)	465-TGACGTGT-458		500-AGC- 498 -- 7-- 490-CAGT-487	529-AATTCAAG-522		598-GCTGAC-593
	3_21	1019	607-ACTCATA- 601		638-ACTCATA-632				651-TGACGTGT- 644	662-AGTAATAT- 669	696-AGC- 694 -- 7-- 686-CAGT- 683	725-AATTCAAG-718		
8_2	**0_32**	560		334-CCGTCC-339 (3)									453-TTGTACTA-460 (2)	
	1_34	823	360-ACTCATA-354		391-ACTCATA-385	405-AATTCAAG-398	459-CACGTC- 464 (3)	464-GACGTG- 459 (2)	465-TGACGTGT-458		500-AGC- 498 -- 7-- 490-CAGT-487	529-AATTCAAG-522		598-GCTGAC-593
	1_45	823	360-ACTCATA-354		391-ACTCATA-385	405-AATTCAAG-398	459-CACGTC- 464 (3)	464-GACGTG- 459 (2)	465-TGACGTGT-458		500-AGC- 498 -- 7-- 490-CAGT-487	529-AATTCAAG-522		598-GCTGAC-593
18_1	**17**	560		334-CCGTCC-339 (3)									453-TTGTACTA-460 (2)	
19_1	**2_26**	823	350-ACTCATA-344		381-ACTCATA-375	395-AATTCAAG-388	449-CACGTC- 454 (3)	454-GACGTG- 449 (2)	455-TGACGTGT- 448	466-AGTAATAT-473	500-AGC- 498 -- 7-- 490-CAGT-487			598-GCTGAC-593
	2_31	823	360-ACTCATA-354		391-ACTCATA-385	405-AATTCAAG-398	459-CACGTC- 464 (3)	464-GACGTG- 459 (2)	465-TGACGTGT-458		500-AGC- 498 -- 7-- 490-CAGT-487	529-AATTCAAG-522		598-GCTGAC-593
20_2	**5**	832	359-ACTCATA-353		390-ACTCATA-384	404-AATTCAAG-397	458-CACGTC-463 (3)	463-GACGTG-458 (2)	464-TGACGTGT-457	475-AGTAATAT-482		538-AATTCAAG-531		
26_L5.S.R.	**17**	822	349-ACTCATA-343		380-ACTCATA-374	394-AATTCAAG-387	448-CACGTC-453 (3)	453-GACGTG-448 (2)	454-TGACGTGT-447	465-AGTAATAT-472	499-AGC-497 -- 7-- 489- CAGT-486			597-GCTGAC-592
27_2	**47**	560		334-CCGTCC-339 (3)									453-TTGTACTA-460 (2)	
	56	823	360-ACTCATA-354		391-ACTCATA-385	405-AATTCAAG-398	459-CACGTC- 464 (3)	464-GACGTG- 459 (2)	465-TGACGTGT-458		500-AGC- 498 -- 7-- 490-CAGT-487	529-AATTCAAG-522		598-GCTGAC-593
	57	823	350-ACTCATA-344		381-ACTCATA-375	395-AATTCAAG-388	449-CACGTC- 454 (3)	454-GACGTG- 449 (2)	455-TGACGTGT- 448	466-AGTAATAT-473	500-AGC- 498 -- 7-- 490-CAGT-487			598-GCTGAC-593
			27	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**

ScanWM-P

Cell line	Clone	Length (bp)	RE: GSN, hor1 /BF: BLZ1, bZIP	RE: PAT2 /BF: BBBF	RE: TGA1	RE: W /BF: WRKY	RE: AuxRE	RE: Box III	RE: RY /BF: ABI3
2_4	**57**	1019				801-TATGACT-807		969-TCAGTTAATG-978	982-CATGCA-977
4_5	**1_16**	823	499-GCGAGTCAT-491		596-TGACGGGC-589	605-TATGACT-611			786-CATGCA-781
	3_21	1019				801-TATGACT-807		969-TCAGTTAATG-978	982-CATGCA-977
8_2	**0_32**	560					537-ATGTCCCCT-545	530-TCAGTTAATG-539	
	1_34	823	499-GCGAGTCAT-491		596-TGACGGGC-589	605-TATGACT-611			786-CATGCA-781
	1_45	823	499-GCGAGTCAT-491		596-TGACGGGC-589				786-CATGCA-781
18_1	**17**	560					537-ATGTCCCCT-545	530-TCAGTTAATG-539	
19_1	**2_26**	823			596-TGACGGGC- 589	605-TATGACT- 611			786-CATGCA-781
	2_31	823	499-GCGAGTCAT-491		596-TGACGGGC-589	605-TATGACT-611			786-CATGCA-781
20_2	**5**	832		523-CTCAAAAAATGTAAATCGAATCA-501		614-TATGACT- 620		782-TCAGTTAATG-791	795-CATGCA-790
26_L5.S.R.	**17**	822			595-TGACGGGC-588	604-TATGACT- 610			785-CATGCA-780
27_2	**47**	560					537-ATGTCCCCT-545	530-TCAGTTAATG-539	
	56	823	499-GCGAGTCAT-491		596-TGACGGGC-589	605-TATGACT-611			786-CATGCA-781
	57	823			596-TGACGGGC- 589	605-TATGACT- 611			786-CATGCA-781
			27	**28**	**29**	**30**	**31**	**32**	**33**

D

AOX1 intron 1

Nsite-PL

Cell line	Clone	Lenght (bp)	RE: GTAC3 /BF: CRR1; SBP	RE: GTAC2 /BF: CRR1 RE: GTAC3 /BF: CRR1, SBP (2)	RE: AE2 /BF: AEF	RE: GA1 /BF: BPC1	RE: PY /BF: DOF3	RE: I-box	RE: Cp /BF: Cp RE: W2 /BF: WRKY53 (9)	RE: Site II	RE: MYC /BF: MYC	RE: C2a /BF: root factor	RE: p33TCP20 /BF: p33TCP20	RE: TGGGCC/T/BF: IPA1 + PCF1/2 RE: TGGGCC/T /BF: TB1 RE: Site II RE: Site II /BF: TCP	RE: Site II	RE: MYC2/3/4 /BF: MYC2/3/4
2_4	**50**	630		252-TGTACC-257 (3)		322-AGAAAGAAA-314	326-AAAAAGAAA-318	385-GATAAGA-391			452-CAGCTGA-446					
	51	712			402-AGAAACAA-395	406-AGAAAGAAA-398	410-AAAAAGAAA-402					475-GCTTATC-469			514-TGGGTC-509	
4_5	**1**	712			402-AGAAACAA-395	406-AGAAAGAAA-398	410-AAAAAGAAA-402					475-GCTTATC-469			514-TGGGTC-509	
8_2	**4**	707			397-AGAAACAA-390	401-AGAAAGAAA-393	405-AAAAAGAAA-397					470-GCTTATC-464			509-TGGGTC-504	
18_1	**1**	636	222-TGTACC-217 (3)	265-TGTACC-270 (3)	324-AGAAACAA-317	328-AGAAAGAAA-320	332-AAAAAGAAA-324		431-TTGACC-436 (13)	438-TGGGTC-433		399-GCTTATC-393				516-CACGAG-511
19_1	**4**	707	281-TGTACC-276 (3)	333-TGTACC-338 (3)	396-AGAAACAA-389							465-GCTTATC-459	500-GCCCA-504	504-TGGGCC-499 (4)		
20_2	**1**	707	281-TGTACC-276 (3)	333-TGTACC-338 (3)	396-AGAAACAA-389							465-GCTTATC-459	500-GCCCA-504	504-TGGGCC-499 (4)		
26_L5.S.R.	**1**	712			402-AGAAACAA-395	406-AGAAAGAAA-398	410-AAAAAGAAA-402					475-GCTTATC-469			514-TGGGTC-509	
27_2	**1**	707			397-AGAAACAA-390		405-AAAAAGAAA-397					470-GCTTATC-464			509-TGGGTC-504	
	47	700			390-AGAAACAA-383	394-AGAAAGAAA-386	398-AAAAAGAAA-390					463-GCTTATC-457				
	52	707			397-AGAAACAA-390	401-AGAAAGAAA-393	405-AAAAAGAAA-397					470-GCTTATC-464			509-TGGGTC-504	
			39	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**

ScanWM-P

Cell line	Clone	Lenght (bp)
2_4	**50**	630
	51	712
4_5	**1**	712
8_2	**4**	707
18_1	**1**	636
19_1	**4**	707
20_2	**1**	707
26_L5.S.R.	**1**	712
27_2	**1**	707
	47	700
	52	707

E

AOX2a intron 1

Nsite-PL

Cell line	Clone	Lenght (bp)
2_4	**47**	2071
4_5	**8**	2070
8_2	**3**	2070
	17	2068
18_1	**S1**	2076
	L56	1981
	L146	1957
	L177	1979
	L214	1980
	L219	1957
19_1	**4**	1987
	14	1989
	64	1985
20_2	**3**	1987
	31	1984
	35	1985
26_L5.S.R.	**S5**	2070
	L13	2070
	L101	2070
	L128	2061
	L142	2079
27_2	**27**	2070

ScanWM-P

Cell line	Clone	Lenght (bp)	RE: MYC	RE: C1	RE: MYC	RE: MYC /BF: MYC	RE: DLEC2A /BF: MAT2 RE: G	RE: ABRE A /BF: ABF (2)	RE: ABRE2	RE: ABRE2	RE: ABRE	RE: Em1b /BF: EmBP1	RE: G /BF: RITA1, bZIP RE: ABRE1	RE: CGACG	RE: unknown	RE: W /BF: WRKY
2_4	**47**	2071			1340-CATGTG-1345	1345-CACATG-1340	1465-GCCACGTCGTAT-1476 (2)	1465-GCCACGTCGT-1474 (2)		1467-CACGTC-1472	1471-ACGTG-1467	1471-ACGTGGCTA-1463	1472-GACGTG-1467 (2)	1473-CGACG-1469	1474-ACGACGTGGC-1465 (2)	
4_5	**8**	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
8_2	**3**	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
	17	2068			1337-CATGTG-1342	1342-CACATG-1337	1462-GCCACGTCGTAT-1473 (2)	1462-GCCACGTCGT-1471 (2)		1464-CACGTC-1469	1468-ACGTG-1464	1468-ACGTGGCTA-1460	1469-GACGTG-1464 (2)	1470-CGACG-1466	1471-ACGACGTGGC-1462 (2)	
18_1	**S1**	2076			1363-CATGTG-1368	1368-CACATG-1363	1470-GCCACGTCGTAT-1481 (2)	1470-GCCACGTCGT-1479 (2)		1472-CACGTC-1477	1476-ACGTG-1472	1476-ACGTGGCTA-1468	1477-GACGTG-1472 (2)	1478-CGACG-1474	1479-ACGACGTGGC-1470 (2)	
	L56	1981					1375-GCCACGTCGTAT-1386 (2)	1375-GCCACGTCGT-1384 (2)		1377-CACGTC-1382	1381-ACGTG-1377	1381-ACGTGGCTA-1373	1382-GACGTG-1377 (2)	1383-CGACG-1379	1384-ACGACGTGGC-1375 (2)	
	L146	1957					1351-GCCACGTCGTAT-1362 (2)	1351-GCCACGTCGT-1360 (2)		1353-CACGTC-1358	1357-ACGTG-1353	1357-ACGTGGCTA-1349	1358-GACGTG-1353 (2)	1359-CGACG-1355	1360-ACGACGTGGC-1351 (2)	
	L177	1979					1373-GCCACGTCGTAT-1384 (2)	1373-GCCACGTCGT-1382 (2)		1375-CACGTC-1380	1379-ACGTG-1375	1379-ACGTGGCTA-1371	1380-GACGTG-1375 (2)	1381-CGACG-1377	1382-ACGACGTGGC-1373 (2)	
	L214	1980					1374-GCCACGTCGTAT-1385 (2)	1374-GCCACGTCGT-1383 (2)		1376-CACGTC-1381	1380-ACGTG-1376	1380-ACGTGGCTA-1372	1381-GACGTG-1376 (2)	1382-CGACG-1378	1383-ACGACGTGGC-1374 (2)	
	L219	1957					1351-GCCACGTCGTAT-1362 (2)	1351-GCCACGTCGT-1360 (2)		1353-CACGTC-1358	1357-ACGTG-1353	1357-ACGTGGCTA-1349	1358-GACGTG-1353 (2)	1359-CGACG-1355	1360-ACGACGTGGC-1351 (2)	
19_1	**4**	1987	1253-CATGTG-1258		1258-CACATG-1253											1389-TATGACC-1383
	14	1989	1255-CATGTG-1260		1260-CACATG-1255											1391-TATGACC-1385
	64	1985	1251-CATGTG-1256		1256-CACATG-1251											1387-TATGACC-1381
20_2	**3**	1987	1253-CATGTG-1258		1258-CACATG-1253											1389-TATGACC-1383
	31	1984	1250-CATGTG-1255		1255-CACATG-1250											1386-TATGACC-1380
	35	1985	1251-CATGTG-1256		1256-CACATG-1251											1387-TATGACC-1381
26_L5.S.R.	**S5**	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
	L13	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
	L101	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
	L128	2061			1330-CATGTG-1335	1335-CACATG-1330	1455-GCCACGTCGTAT-1466 (2)	1455-GCCACGTCGT-1464 (2)		1457-CACGTC-1462	1461-ACGTG-1457	1461-ACGTGGCTA-1453	1462-GACGTG-1457 (2)	1463-CGACG-1459	1464-ACGACGTGGC-1455 (2)	
	L142	2079		1300-GAAAAAAAAAAAG-1287	1348-CATGTG-1353	1353-CACATG-1348	1473-GCCACGTCGTAT-1484 (2)	1473-GCCACGTCGT-1482 (2)	1475-CACGTC-1480	1483-TACGAC1478	1479-ACGTG-1475	1479-ACGTGGCTA-1471	1480-GACGTG-1475 (2)	1481-CGACG-1477	1482-ACGACGTGGC-1473 (2)	
27_2	**27**	2070			1339-CATGTG-1344	1344-CACATG-1339	1464-GCCACGTCGTAT-1475 (2)	1464-GCCACGTCGT-1473 (2)		1466-CACGTC-1471	1470-ACGTG-1466	1470-ACGTGGCTA-1462	1471-GACGTG-1466 (2)	1472-CGACG-1468	1473-ACGACGTGGC-1464 (2)	
			39	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**

F

AOX2b intron 1

Nsite-PL

Cell line	Clone	Length (bp)	RE: WUS2 /BF: WUS	RE: AC-I /BF: MYB1/B4
2_4	**57**	1019	796-AGTAATAT-803	
4_5	**1_16**	823	600-AGTAATAT- 607	
	3_21	1019	796-AGTAATAT-803	
8_2	**0_32**	560		
	1_34	823	600-AGTAATAT- 607	
	1_45	823	600-AGTAATAT- 607	
18_1	**17**	560		
19_1	**2_26**	823	600-AGTAATAT-607	798-ACCTATC- 804
	2_31	823	600-AGTAATAT- 607	
20_2	**5**	832	609-AGTAATAT- 616	807-ACCTATC-813
26_L5.S.R.	**17**	822	599-AGTAATAT-606	797-ACCTATC-803
27_2	**47**	560		
	56	823	600-AGTAATAT- 607	
	57	823	600-AGTAATAT- 607	798-ACCTATC- 804
			39	**40**

ScanWM-P

Cell line	Clone	Length (bp)
2_4	**57**	1019
4_5	**1_16**	823
	3_21	1019
8_2	**0_32**	560
	1_34	823
	1_45	823
18_1	**17**	560
19_1	**2_26**	823
	2_31	823
20_2	**5**	832
26_L5.S.R.	**17**	822
27_2	**47**	560
	56	823
	57	823

D

AOX1 intron 1

Nsite-PL

Cell line	Clone	Lenght (bp)	RE: W3 /BF: MINI3	RE: GATA /BF: R-GATA	RE: Cp BS /BF: Cp RE: W2 /BF: WRKY53 (9) RE: W /BF: WRKY	RE: W1/BF: MINI3 RE:W /BF: WRKY33	RE: CARE H (2) RE: CARE	RE: E /BF: bHLH122 (3)	RE: E /BF: CIB2/4/5 RE: E /BF: CIB1; CIB1/3; CIB1/4; CIB2/4; CIB2/5; CIB4/5	RE: MYC2/3/4-3 /BF: MYC2/3/4	RE: AC-I /BF: MYB61
2_4	**50**	630	537-TTGACAA-543	542-AATAGGATA-550							
	51	712								592-CACGAG-587	
4_5	**1**	712								592-CACGAG-587	
8_2	**4**	707								587-CACGAG-582	617-ACCATA-622
18_1	**1**	636			543-TTGACC-548 (13)	543-TTGACCA-549 (2)					546-ACCATA-551
19_1	**4**	707					551-GTCATC-546 (3)	551-CACTTG-556 (3)	556-CAAGTG-551 (2)	585-CACGAG-580	
20_2	**1**	707					551-GTCATC-546 (3)	551-CACTTG-556 (3)	556-CAAGTG-551 (2)	585-CACGAG-580	
26_L5.S.R.	**1**	712								592-CACGAG-587	
27_2	**1**	707								587-CACGAG-582	617-ACCATA-622
	47	700								580-CACGAG-575	610-ACCATA-615
	52	707								587-CACGAG-582	617-ACCATA-622
			53	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**

ScanWM-P

Cell line	Clone	Lenght (bp)
2_4	**50**	630
	51	712
4_5	**1**	712
8_2	**4**	707
18_1	**1**	636
19_1	**4**	707
20_2	**1**	707
26_L5.S.R.	**1**	712
27_2	**1**	707
	47	700
	52	707

E

AOX2a intron 1

Nsite-PL

Cell line	Clone	Lenght (bp)
2_4	**47**	2071
4_5	**8**	2070
8_2	**3**	2070
	17	2068
18_1	**S1**	2076
	L56	1981
	L146	1957
	L177	1979
	L214	1980
	L219	1957
19_1	**4**	1987
	14	1989
	64	1985
20_2	**3**	1987
	31	1984
	35	1985
26_L5.S.R.	**S5**	2070
	L13	2070
	L101	2070
	L128	2061
	L142	2079
27_2	**27**	2070

ScanWM-P

Cell line	Clone	Lenght (bp)	RE: A1 /BF: CBF1	RE: W /BF: WRKY	RE: AT1 /BF: AT1	RE: AT1 /BF: AT1	RE: RY /BF: ABI3	RE: MYC /BF: MYC	RE: MYC	RE: MYC	RE: CRE	RE: ATCATC	RE: TGA1	RE: AGL2 /BF: AGL2	RE: Box 1 /BF: TFHP1	RE: ABRE1/3 /BF: ABI5
2_4	**47**	2071	1475-TACGAC-1470			1533-ATTAGTTTTAAT-1522	1669-CATGCA-1674	1673-CACATG-1678	1678-CATGTG-1673		1757-CAAGATCATC-1766	1761-ATCATC-1766				
4_5	**8**	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
8_2	**3**	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
	17	2068	1472-TACGAC-1467			1530-ATTAGTTTTAAT-1519	1666-CATGCA-1671	1670-CACATG-1675	1675-CATGTG-1670		1754-CAAGATCATC-1763	1758-ATCATC-1763				
18_1	**S1**	2076	1480-TACGAC-1475			1538-ATTAGTTTTAAT-1527	1674-CATGCA-1679	1678-CACATG-1683	1683-CATGTG-1678		1762-CAAGATCATC-1771	1766-ATCATC-1771				
	L56	1981	1385-TACGAC-1380			1443-ATTAGTTTTAAT-1432	1579-CATGCA-1584	1583-CACATG-1588	1588-CATGTG-1583	1629-CATGTG-1634	1667-CAAGATCATC-1676	1671-ATCATC-1676				
	L146	1957	1361-TACGAC-1356			1419-ATTAGTTTTAAT-1408	1555-CATGCA-1560	1559-CACATG-1564	1564-CATGTG-1559		1643-CAAGATCATC-1652	1647-ATCATC-1652				
	L177	1979	1383-TACGAC-1378			1441-ATTAGTTTTAAT-1430	1577-CATGCA-1582	1581-CACATG-1586	1586-CATGTG-1581		1665-CAAGATCATC-1674	1669-ATCATC-1674				
	L214	1980	1384-TACGAC-1379			1442-ATTAGTTTTAAT-1431	1578-CATGCA-1583	1582-CACATG-1587	1587-CATGTG-1582		1666-CAAGATCATC-1675	1670-ATCATC-1675				
	L219	1957	1361-TACGAC-1356			1419-ATTAGTTTTAAT-1408	1555-CATGCA-1560	1559-CACATG-1564	1564-CATGTG-1559		1643-CAAGATCATC-1652	1647-ATCATC-1652				
19_1	**4**	1987		1411-GATGACC-1405	1448-ATTAATTTTAAT-1437								1686-TGACGATC-1679	1850-AGCCATAATTGCATA-1836	1862-CATCCACGTGCACT-1875	1864-TCCACGTGCA-1873
	14	1989		1413-GATGACC-1407	1450-ATTAATTTTAAT-1439								1688-TGACGATC-1681	1852-AGCCATAATTGCATA-1838	1864-CATCCACGTGCACT-1877	1866-TCCACGTGCA-1875
	64	1985		1409-GATGACC-1403	1446-ATTAATTTTAAT-1435								1684-TGACGATC-1677	1848-AGCCATAATTGCATA-1834	1860-CATCCACGTGCACT-1873	1862-TCCACGTGCA-1871
20_2	**3**	1987		1411-GATGACC-1405	1448-ATTAATTTTAAT-1437								1686-TGACGATC-1679	1850-AGCCATAATTGCATA-1836		1864-TCCACGTGCA-1873
	31	1984		1408-GATGACC-1402	1445-ATTAATTTTAAT-1434								1683-TGACGATC-1676	1847-AGCCATAATTGCATA-1833	1859-CATCCACGTGCACT-1872	1861-TCCACGTGCA-1870
	35	1985		1409-GATGACC-1403	1446-ATTAATTTTAAT-1435								1684-TGACGATC-1677	1848-AGCCATAATTGCATA-1834	1860-CATCCACGTGCACT-1873	1862-TCCACGTGCA-1871
26_L5.S.R.	**S5**	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
	L13	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
	L101	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
	L128	2061	1465-TACGAC-1460			1523-ATTAGTTTTAAT-1512		1663-CACATG-1668	1668-CATGTG-1663		1747-CAAGATCATC-1756	1751-ATCATC-1756				
	L142	2079	1483-TACGAC1478			1541-ATTAGTTTTAAT-1530		1681-CACATG-1686	1686-CATGTG-1681		1765-CAAGATCATC-1774	1769-ATCATC-1774				
27_2	**27**	2070	1474-TACGAC-1469			1532-ATTAGTTTTAAT-1521	1668-CATGCA-1673	1672-CACATG-1677	1677-CATGTG-1672		1756-CAAGATCATC-1765	1760-ATCATC-1765				
			53	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**

F

AOX2b intron 1

Nsite-PL

Cell line	Clone	Length (bp)
2_4	**57**	1019
4_5	**1_16**	823
	3_21	1019
8_2	**0_32**	560
	1_34	823
	1_45	823
18_1	**17**	560
19_1	**2_26**	823
	2_31	823
20_2	**5**	832
26_L5.S.R.	**17**	822
27_2	**47**	560
	56	823
	57	823

ScanWM-P

Cell line	Clone	Length (bp)
2_4	**57**	1019
4_5	**1_16**	823
	3_21	1019
8_2	**0_32**	560
	1_34	823
	1_45	823
18_1	**17**	560
19_1	**2_26**	823
	2_31	823
20_2	**5**	832
26_L5.S.R.	**17**	822
27_2	**47**	560
	56	823
	57	823

D

AOX1 intron 1

NsiteI-PL

Cell line	Clone	Lenght (bp)
2_4	50	630
	51	712
4_5	1	712
8_2	4	707
18_1	1	636
19_1	4	707
20_2	1	707
26_L5.S.R.	1	712
27_2	1	707
	47	700
	52	707

ScaniWM-P

Cell line	Clone	Lenght (bp)
2_4	50	630
	51	712
4_5	1	712
8_2	4	707
18_1	1	636
19_1	4	707
20_2	1	707
26_L5.S.R.	1	712
27_2	1	707
	47	700
	52	707

E

AOX2a intron 1

NsiteI-PL

Cell line	Clone	Lenght (bp)
2_4	47	2071
4_5	8	2070
8_2	3	2070
	17	2068
18_1	S1	2076
	L56	1981
	L146	1957
	L177	1979
	L214	1980
	L219	1957
19_1	4	1987
	14	1989
	64	1985
20_2	3	1987
	31	1984
	35	1985
26_L5.S.R.	S5	2070
	L13	2070
	L101	2070
	L128	2061
	L142	2079
27_2	27	2070

ScaniWM-P

Cell line	Clone	Lenght (bp)	RE: G /BF: PG1	RE: E4 /BF: DBPF1/2 RE: G (3) RE: G /BF: RITA1, bZIP (2) RE: G /BF: TAF1, ABI3, PG1	RE: ABRE	RE: Em1b /BF: EmBP1	RE: Em1b /BF: EmBP1	RE: ABRE	RE: Em1b /BF: EmBP1	RE: E4 /BF: DBPF1/2 RE: G (3) RE: G /BF: RITA1, bZIP (2) RE: G /BF: TAF1, ABI3, PG1	RE: G /BF: PG1	RE: ABRE /BF: ABI5	RE: E3 /BF: DBPF1/2	RE: Box 1 /BF: TFHP1	RE: CRE
2_4	47	2071											1949-CACGCG-1954		
4_5	8	2070											1948-CACGCG-1953		
8_2	3	2070											1948-CACGCG-1953		
	17	2068											1946-CACGCG-1951		
18_1	S1	2076											1954-CACGCG-1959		
	L56	1981											1859-CACGCG-1864		1966-CAAGGGAATC-1957
	L146	1957											1835-CACGCG-1840		1942-CAAGGGAATC-1933
	L177	1979											1857-CACGCG-1862		1964-CAAGGGAATC-1955
	L214	1980											1858-CACGCG-1863		1965-CAAGGGAATC-1956
	L219	1957											1835-CACGCG-1840		1942-CAAGGGAATC-1933
19_1	4	1987	1865-CCACGTGCACTCT-1877	1866-CACGTG-1871 (9)	1867-ACGTG-1871	1867-ACGTGCACT-1875		1870-ACGTG-1866	1870-ACGTGGATG-1862	1871-CACGTG-1866 (9)	1872-GCACGTGGATGAG-1860	1873-TGCACGTGGA-1864		1875-AGTGCACGTGGATG-1862	
	14	1989	1867-CCACGTGCACTCT-1879	1868-CACGTG-1873 (9)	1869-ACGTG-1873		1869-ACGTGACAG-1877	1872-ACGTG-1868	1872-ACGTGGATG-1864	1873-CACGTG-1868 (9)	1874-GCACGTGGATGAG-1862	1875-TGCACGTGGA-1866		1877-AGTGCACGTGGATG-1864	
	64	1985	1863-CCACGTGCACTCT-1875	1864-CACGTG-1869 (9)	1865-ACGTG-1869	1865-ACGTGCACT-1873		1868-ACGTG-1864	1868-ACGTGGATG-1860	1869-CACGTG-1864 (9)	1870-GCACGTGGATGAG-1858	1871-TGCACGTGGA-1862		1873-AGTGCACGTGGATG-1860	
20_2	3	1987		1866-CACGTG-1871 (9)	1867-ACGTG-1871	1867-ACGTGCACT-1875		1870-ACGTG-1866	1870-ACGTGGATG-1862	1871-CACGTG-1866 (9)		1873-TGCACGTGGA-1864			
	31	1984	1862-CCACGTGCACTCT-1874	1863-CACGTG-1868 (9)	1864-ACGTG-1868	1864-ACGTGCACT-1872		1867-ACGTG-1863	1867-ACGTGGATG-1859	1868-CACGTG-1863 (9)	1869-GCACGTGGATGAG-1857	1870-TGCACGTGGA-1861		1872-AGTGCACGTGGATG-1859	
	35	1985	1863-CCACGTGCACTCT-1875	1864-CACGTG-1869 (9)	1865-ACGTG-1869	1865-ACGTGCACT-1873		1868-ACGTG-1864	1868-ACGTGGATG-1860	1869-CACGTG-1864 (9)	1870-GCACGTGGATGAG-1858	1871-TGCACGTGGA-1862		1873-AGTGCACGTGGATG-1860	
26_L5.S.R.	S5	2070											1948-CACGCG-1953		
	L13	2070											1948-CACGCG-1953		
	L101	2070											1948-CACGCG-1953		
	L128	2061											1939-CACGCG-1944		
	L142	2079											1957-CACGCG-1962		
27_2	27	2070											1948-CACGCG-1953		
			67	68	69	70	71	72	73	74	75	76	77	78	79

F

AOX2b intron 1

NsiteI-PL

Cell line	Clone	Length (bp)
2_4	57	1019
4_5	1_16	823
	3_21	1019
8_2	0_32	560
	1_34	823
	1_45	823
18_1	17	560
19_1	2_26	823
	2_31	823
20_2	5	832
26_L5.S.R.	17	822
27_2	47	560
	56	823
	57	823

ScaniWM-P

Cell line	Clone	Length (bp)
2_4	57	1019
4_5	1_16	823
	3_21	1019
8_2	0_32	560
	1_34	823
	1_45	823
18_1	17	560
19_1	2_26	823
	2_31	823
20_2	5	832
26_L5.S.R.	17	822
27_2	47	560
	56	823
	57	823