



# Heterogenity of genes encoding the low molecular weight glutenin subunits of bread wheat

T. Chardot, T. Do, Lise Perret, M. Laurière

## ► To cite this version:

T. Chardot, T. Do, Lise Perret, M. Laurière. Heterogenity of genes encoding the low molecular weight glutenin subunits of bread wheat. Workshop on Plant Biopolymer Science Location NANTES, FRANCE, Jun 2001, Nantes, France. pp.24-30, <10.1039/9781847551672-00024>. <hal-04701742>

**HAL Id: hal-04701742**

**<https://hal.science/hal-04701742v1>**

Submitted on 27 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Heterogeneity of genes encoding the low molecular weight glutenin subunits of bread wheat.

T. Chardot, T. Do, L. Perret and M Laurière

UMR de Chimie Biologique, INRA INA-PG Centre de Biotechnologie Agro-Industrielle,  
F- 78850 Thiverval-Grignon.

## 1 ABSTRACT

Among the wheat storage proteins, the low molecular weight glutenin subunit (LMW-GS) group is probably the more complex. It encompasses several subgroups which differ by their N-terminal amino acid sequences. To get more information on the diversity of the genes encoding the LMW-GS, we attempted to clone some of them. Two different strategies were used to amplify their coding sequences from genomic DNA using PCR techniques.

-The first strategy used primers corresponding to the N- and C-terminal sequences of already published mature peptides. Several DNA fragments with various sizes were amplified and more than 2500 clones were obtained. Among them 95 clones were further analyzed. The size of their insert ranged from 1200 to 800 and less bp. Restriction analysis of these inserts showed that they can be gathered in ten families. 13 clones have been sequenced, all exhibited the features of LMW-GS coding sequences. For 8 of the sequences, the coding frame was interrupted by 1 to 10 stops codons, for 3 sequences only parts of a LMW-GS coding frame could be recognized. The two last sequences were encoding for LMW-GS which were 270 and 327 residues long.

–The second strategy used primers anchored in the promoters, and in the C-terminal part. We obtained 4 different clones, with complete coding sequence, and part of the promoting sequence. They were LMW-GS coding sequences (852 bp, 284 amino acid residues), and did not contain any interruption of the coding frame.

The use of PCR technique is a convenient way for investigating the complexity of the wheat LMW-GS multigenic family. The existence of numerous pseudo genes gives interesting insights into the evolution of this multigenic family. Extinguished LMW-GS may be a reservoir for future ?

This work was supported by a grant from INRA (Action Nouvelle Soutenue to TC and ML).

## 2 INTRODUCTION

The sequence of the wheat genome is still unknown. Its size, much larger than that of the human genome, is a roadblock to the study of the organization of the genes. The knowledge of that organization would be useful from agronomic, nutritional and health points of view. Concerning the genes encoding the storage proteins expressed constitutively in the grain, few is known about their precise number, their heterogeneity and their organization at the molecular level. They are encoded by clusters of homologous genes on the homeologous chromosomes 1 and 6. Among them, those encoding the low molecular weight glutenin subunits (LMW-GS) are probably the more heterogeneous and the less known. As other prolamins they are secretory proteins with a signal peptide of 20 amino acids. Depending of the N-terminal sequence of their mature peptide, several subgroups were identified. There are the s- and m-LMW-GS, whether they have serine or methionine, as N-terminal or the LMW-GS with sequences homologous to alpha/beta- and gamma-gliadins. The s-LMW-GS appear heterogeneous and highly expressed in the endosperm. Only two complete sequences were described [1, 2]. Using different primer sets, we tried to clone these genes using PCR. DNA fragments were cloned, sequenced and analysed. We discuss here the heterogeneity of the sequences we got.

## 3 MATERIAL AND METHODS

### 3.1 Genomic DNA preparation

DNA was prepared according to Benito et al (1993) [3] from 250 mg of etiolated seedling, 3 days old, of cv Nepawa. It was stored at -20°C until used.

### 3.2 DNA amplification and cloning:

LMW-GS primers for PCR were synthesized by MWG biotech (Germany). They are described in table I. PCR reaction (25 µL final volume) contained Taq polymerase, 100 pmoles of the oligonucleotides primers, 2500 pmoles dNTP, *T aestivum* genomic DNA (20 to 100 ng). 25 to 30 cycles were performed. A 1 min denaturation step at 94°C was followed by a hybridization step at 60 or 62°C for 0.5 to 2 min, followed by a 1.5 min elongation step at 72°C. The experiment was terminated by a 4 min step at 72 °C. The PCR fragments were purified using Wizard PCR prep kit, either directly, or after electrophoresis separation, and

were cloned in the PGEM-T vector. The plasmidic DNA were first restriction analyzed, and then sequenced by MWG Biotech (Germany).

### **3.3 Sequence alignments:**

Sequences were analyzed using the BioEdit software V 4.7.1 by Tom Hall. The Blosum62 matrix was used for sequences alignments.

## **4 RESULTS**

Figure 1 summarizes the strategy used to amplify LMW-GS coding and non coding sequences and the length of the clones studied.

### **4.1 Amplification of DNA sequences, using primers anchored at the N- and C-terminal sequences of mature s-LMWG-GS**

Numerous DNA fragments with various sizes were amplified (not shown) using oligonucleotide primers LMW004-1 / LMW002 and - LMW006-1 / LMW006-2 designed from Masci *et al.* [2] They have been cloned into the pGEM-T vector. More than 2500 clones were obtained, which were white blue screened. Among the positive clones, 95 were further restriction analyzed using either Bstx I + Sph I, to excise the full DNA fragment inserted into the pGEM-T plasmid, or Nde I + Xho I, to excise clones containing XhoI site. This allowed to distinguish ten different families, according to their restriction maps. The DNA of 13 clones, representing the different families, were sequenced. All the sequences obtained exhibited the features of LMW-GS coding sequences as deduced from blast searches, or from translations. 8 sequences had coding frame interrupted by 1 to 10 stops codons. For 3 sequences, only parts of a LMW-GS coding frame could be recognized from the sequence. Figure 2, shows the alignment of the amino acid sequences of the mature peptide derived from the longer DNA sequences. Clone 208 exhibited the longer size with 327 residues. Clones TD4 and TD6 were respectively 284 and 281 residues long. Clone 321 (270 residues) is shown because it contained only one stop codon. Their N-terminal acid sequence (SHIP), imposed by the primers used, make them putative s-type LMW- GS. They are highly homologous with highly conserved N-terminal (over 13 residues) and C-terminal (over 54 residues). The central domain which exhibited

repetitive stretches rich in glutamine and proline, ranged from 196 (clone 321) to 250 residues (clone 208). It can be divided according to Cassidi *et al.* [4] in three sub domains II to IV. Domain II and IV are mainly responsible for the differences of the size of the clones. Domain III as limited stretches of glutamine and is strongly conserved. All clones contained the 6 conserved cysteines (yellow stars in domain III and V), which are assumed to be engaged in disulfide bridges. Only one additional cysteine was found at a position that can slightly differ in domain IV (blue stars), and which may have a role in polymer formation [5].

## **4.2 Amplification of DNA sequences, using primers anchored at the promoter and C-terminal sequences of LMWG-GS.**

Using primers anchored in the promoters, and in the C-terminal part of LMW-GS, we amplified, using PCR technique, DNA molecules with size ranging from 600 to # 1200 bp, as deduced from agarose gel electrophoresis. The major 1200 bp fragment, was cloned into the pGEM-T plasmid. 240 colonies were obtained, and white/blue screened. 20 colonies were PCR screened, using proper oligonucleotide. All cloned were positives. They exhibited similar sizes, as estimated from 1.5% agarose electrophoresis. Four clones were selected among the positives colonies. They all were containing a # 1350 bp insert. These clones have been sequenced.

### **4.2.1 Analysis of the non coding DNA sequences**

The partial promoting sequences of the four clones, showed strong homologies and typical boxes (results not shown). The endosperm box [6, 7, 8] was located at position –311. It was almost immediately followed by a GCN4 motif (–292), controlling endosperm specific expression [9]. The CAT box was found at position –169. AGGA box was found at –100. The AACA motif was found at position –165. TATA boxes were found at position –62 and –79. Initiator codon was found in the CCACCATGA consensus sequence [8, 10].

#### 4.2.2 Analysis of the coding sequences

All sequences of the previous clones, were coding for a precursor protein, which all belonged to the METRCIP LMW-GS subfamily. They were 284 amino acid residues long, differing only by 3 residues. The four clones sequences exhibited a 20 amino acid residue signal peptide, followed by the N-terminal domain with the conserved 13 residues, a 210 residues repetitive domain. Compared to the above described clones, they differ by repetitive stretches in domain II and by the domain IV which was 12 residues shorter. It shifted the non conserved cysteine found in this last domain. The 54 residues C terminal domain was conserved (Figure 2).

### 5 DISCUSSION

Using N and C terminal directed oligonucleotide sequences in PCR technique, allowed us to amplify and to clone various type of DNA sequences, some of them exhibited typical LMW-GS coding sequences potentially expressed. Others showed various number of stop codons, up to 10 in the typical LMW-GS open reading frame or only partially LMW-GS coding sequences features. That make them putative pseudo genes. Such DNA sequences were already described [11-13]. Their existence is likely be common in different wheat species. They are certainly not expressed *in vivo*, due essentially to the number of stop codons found in the DNA sequences. They may represent reservoirs of variability, that could be revived by mutagenesis. Our goal was to clone s-type LMW-GS which are largely expressed in bread and durum wheat [1, 2] using primers corresponding to the SHIP sequences at the N-terminal. We amplified DNA sequences with SHIP N-terminal coding sequences corresponding to potentially expressed LMW-GS. Due to the PCR technique used, which dictates the 5' and 3' terminal sequences, these DNA sequences were only putative s-LMW-GS. It is interesting to note their high homology to the clone obtained with a primer anchored in the promoter. An other interesting feature is that, with this last strategy we elicited only m-LMW-GS, unlike Masci *et al.* [1, 2]. Presently 31 LMW-GS sequences were described in EMBL. Our work underline again the difficulty to isolate s-type LMW-GS sequences despite the fact that we got more different DNA sequences than Masci *et al.* [1,2], using the same strategy. However with this strategy using primers anchored in the promoter, we amplified DNA sequences with no stop codons in their coding sequences unlike upon use of primers anchored at the N-terminal. Nevertheless PCR technique using both

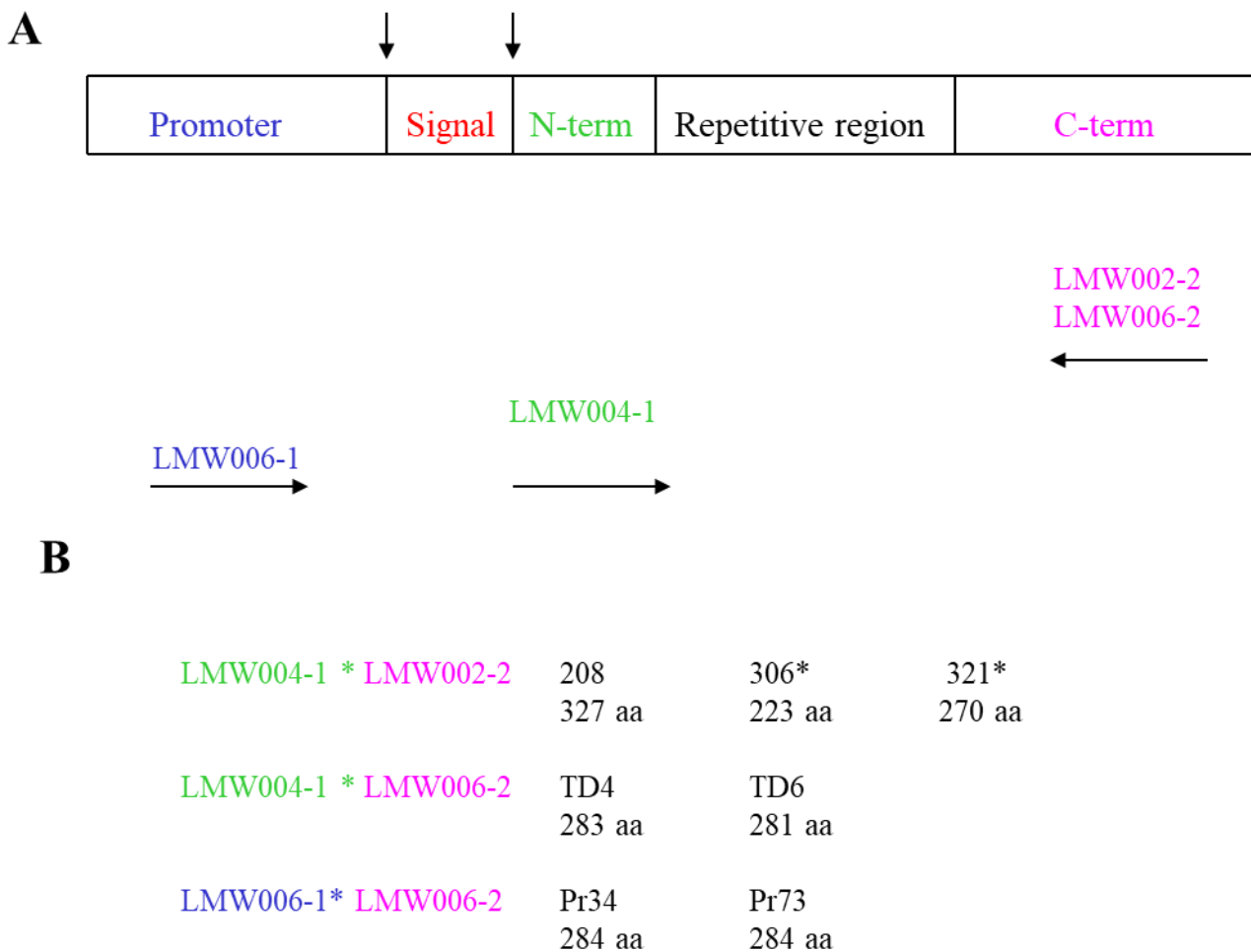
strategy, is an attractive method to investigate the diversity and complexity of wheat reserve protein genes, in the absence of the complete sequencing of the wheat genome.

## 6 FIGURES

**Figure 1:**

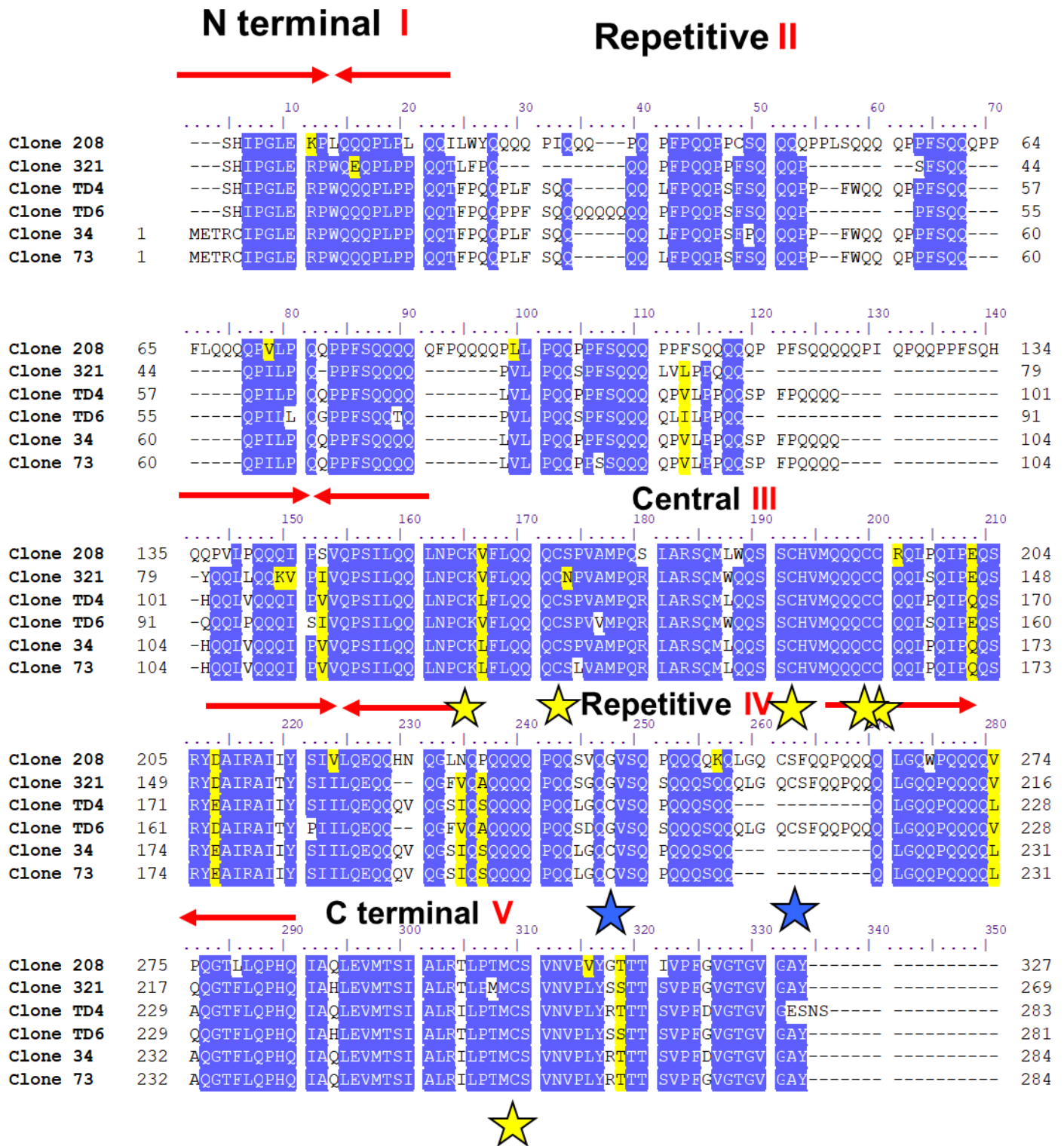
A: Schematic representation of a LMWG gene, with the nucleotidic primers used in this study. Vertical arrows indicate the beginning of the coding frame, and of the mature peptide. Central sub domains were not represented.

B: Summary of clones of interest obtained in this study. \* indicates pseudo genes .



**Figure 2:** DNA fragments obtained upon amplification of genomic DNA from wheat (var Neepawa) with oligonucleotide primers LMW0004-1 x LMW0002-2; LMW0004-1 x LMW0007-2 and LMW0006-1 x

7



**Table I: Oligonucleotides used in this study.**



Primers LMW0004-1; LMW0006-1 LMW0002-2; LMW0006-2 and LMW0007-2 and were aimed at amplifying the sequences coding for LMW-GS using genomic DNA as template.

**LMW004-1** : (Adapted from ref 14)

Nucleic sequence : 5' AGC CAC ATC CCT GGT TTG GA 3'

Proteic translation: Nterm **S H I P** G L → Cterm

**LMW006-1** : (from ref 15)

Nucleic sequence : 5' CGA GCA TAT CCT AAC AGC CCA 3'

**LMW002-2** : (from ref 15)

Nucleic sequence : 3' G TGG CCT CAA CCA CGG ATG 5'

Proteic translation: Nterm ← T G V G A Y Cterm

**LMW006-2** : (from ref 15)

Séquence nucléique : 3' CG TGG CCT CAA CCA CGG ATG 5'

Séquence protéique : N term G T G V G A Y

## Literature

- 1) S. Masci , R. D'Ovidio, D. Lafiandra and D.D. Kasarda. *Theor. Appl. Genet.* 2000, **100**, 396-400
- 2) S. Masci , R. D'Ovidio, D. Lafiandra and D.D. Kasarda,. *Plant physiol.*, 1998 **118** : 1147-1158.
- 3) C. Benito, M. Figueiras, C. Zaragoza, F.J. Gallego and A. de la Peña, *Plant. Mol. Biol* 1993.. **21**, 181-183
- 4) B. G. Cassidy, J. dvorak, O.D. Anderson. *Theor. Appl. Genet.* 1998, **95**, 743-750
- 5) P. R. Shewry and A. S. Tatham. *J. Cereal Sci.* 1997, **25**, 207-227
- 6) J. Forde, J.M. Malpica, N.G. Halford, P.R. Shewry, O.D. Anderson, F.C. Greene, B.J. Mifflin. *Nucleic Acids Res* 1985 **19** , 6817-6832
- 7) M. Kreis, B. G. Forde, S. Rahman B. J Mifflin, P.R. Shewry *J Mol. Biol* 1985 **183** :499-502
- 8) V. Colot, D. Bartels, R. Thompson, R. Flavell. *Mol Gen Genet*, 1989, **216** : 81-90
- 9) Y Onodera, A Suzuki, CY Wu, H Washida, F Takaiwa. *J Biol, Chem*, , 2001, **276** : 14139-14152
- 10) M. Kozak, *Cell*, 1986, **44**: 283-292.
- 11) Tanurdzic, M., Obreht, D. and Vapa, L. J. unpublished
- 12) M. Benmoussa, L. P. Vézina, M. Pagé, S. Yelle, S. Laberge. *Theor. Appl. Genet.* 2000, **100**, 789-793
- 13) R. D'Ovidio, S. Masci, S., C. Marchitelli, P. Tosi, M Simeone, and E. Porceddu. Unpublished
- 14) R. D'Ovidio, C. Marchitelli, L. Ercoli Cardeli and E. Porceddu. *Theor. Appl. Genet.* 1999, **98**, 455-461
- 15) R. D'Ovidio, M Simeone, S. Masci, S. and E. Porceddu. *Theor. Appl. Genet.* 1997, **95**, 1119-1126