the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2008) doi: 10.1111/j.1365-313X.2008.03482.x

TECHNICAL ADVANCE

Generation of a 3D indexed *Petunia* insertion database for reverse genetics

Michiel Vandenbussche[1], Antoine Janssen[2], Jan Zethof[1], Nathalie van Orsouw[2], Janny Peters[1], Michiel J.T. van Eijk[2], Anneke S. Rijpkema[1], Harrie Schneiders[2], Parthasarathy Santhanam[1], Mark de Been[3], Arjen van Tunen[2] and Tom Gerats[1,2,*]

[1]*Radboud University, IWWR/Plant Genetics, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands,*

[2]*Keygene NV, AgroBusiness Park 90, 6708 PW Wageningen, The Netherlands, and*

[3]*Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands*

Received 23 November 2007; revised 1 February 2008; accepted 18 February 2008.

*For correspondence (fax +31 2436 52787; e-mail t.gerats@science.ru.nl).

Summary

BLAST searchable databases containing insertion flanking sequences have revolutionized reverse genetics in plant research. The development of such databases has so far been limited to a small number of model species and normally requires extensive labour input. Here we describe a highly efficient and widely applicable method that we adapted to identify unique transposon-flanking genomic sequences in *Petunia*. The procedure is based on a multi-dimensional pooling strategy for the collection of DNA samples; up to thousands of different templates are amplified from each of the DNA pools separately, and knowledge of their source is safeguarded by the use of pool-specific (sample) identification tags in one of the amplification primers. All products are combined into a single sample that is subsequently used as a template for unidirectional pyrosequencing. Computational analysis of the clustered sequence output allows automatic assignment of sequences to individual DNA sources. We have amplified and analysed transposon-flanking sequences from a *Petunia* transposon insertion library of 1000 individuals. Using 30 DNA isolations, 70 PCR reactions and two GS20 sequencing runs, we were able to allocate around 10 000 transposon flanking sequences to specific plants in the library. These sequences have been organized in a database that can be BLAST-searched for insertions into genes of interest. As a proof of concept, we have performed an *in silico* screen for insertions into members of the *NAM*/*NAC* transcription factor family. All *in silico*-predicted transposon insertions into members of this family could be confirmed *in planta*.

Keywords: insertion mutagenesis, *Petunia*, pyrosequencing, transposon, *dTph1*, NAC family.

Introduction

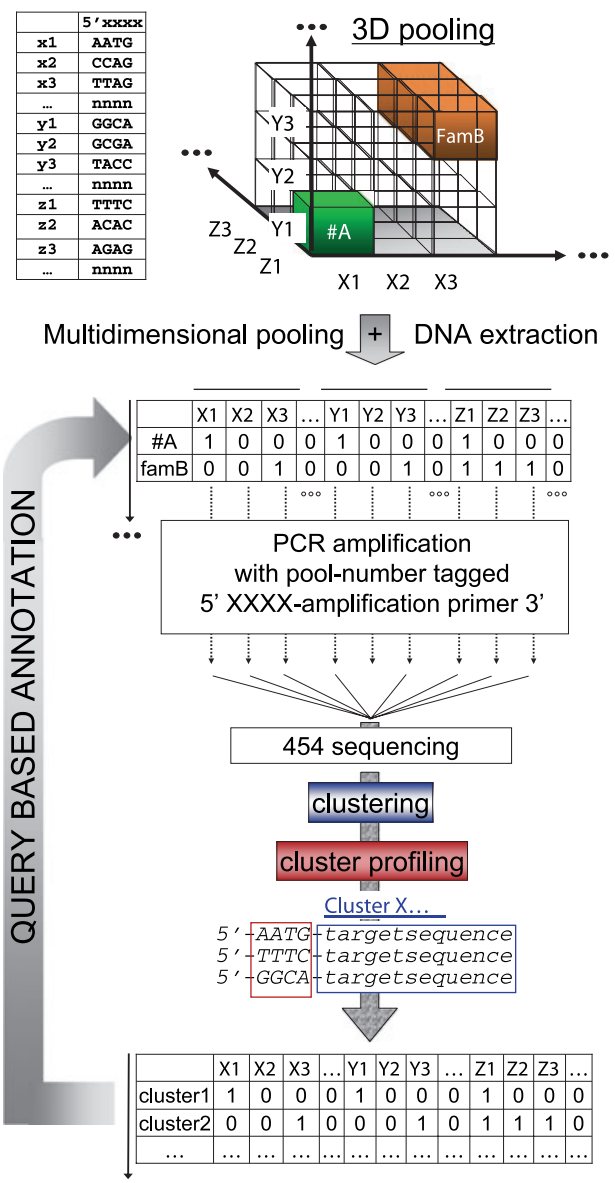
One of the classical challenges in biology is to understand how changes in gene function during evolution have contributed to the diversity of life as it exists today. To study the evolution of gene function, it is necessary to compare loss-of-function mutants for orthologous genes in different species. In Arabidopsis, the publicly available collections of insertion site-sequenced T-DNA lines (Alonso *et al.*, 2003) have revolutionized reverse-genetics approaches, as databases with insertion site flanking sequences can be searched for mutants by a simple sequence homology search instead of having to perform laborious PCR-based assays. Similar platforms are being developed for the monocot species rice and maize (Droc *et al.*, 2006; van Enckevort *et al.*, 2005; Kolesnik *et al.*, 2004; Settles *et al.*, 2007; Yazaki *et al.*, 2004). Plant comparative biology research would greatly benefit if such resources could be developed for a broader range of species. We wish to create such a platform for the model species *Petunia*, taking advantage of the endogenous, highly active *dTph1* transposable element system in line W138 (Gerats *et al.*, 1990). Petunia has several qualities that make

it a fairly amenable model system, amongst which are highly efficient forward and reverse transposon insertion screening methods (Gerats and Vandenbussche, 2005).

As a starting point for further improvements in these methods, we aimed to simultaneously sequence and order *dTph1* flanking sequences amplified from a population of 1000 individuals, by using the powerful 454 liquid-phase massive parallel sequencing technique (Margulies *et al.*, 2005). For reverse-genetics purposes, the origin of every flanking sequence tag in the resulting database needs to be assigned to a specific individual within the population. As 454 sequencing in principle does not allow discrimination between individual DNA sources in a pooled sample, we have developed and applied a general methodology (Figure 1) that allows massive parallel amplification and sequencing of thousands of different fragments that are automatically assigned to their specific origin. We have coupled the use of 5′-encoded PCR primers (see also Binladen *et al.*, 2007; van Orsouw *et al.*, 2007) with a three-dimensional pooling system, and have adapted the original GS20 pyrosequencing protocol (Margulies *et al.*, 2005) to enable unidirectional sequencing of the amplified transposon tag library, followed by automated assignment of sequences to individuals or families within the population by a computational approach.

The three-dimensional (3D) pooling principle (Figure 1) allows identification of a specific individual within a group of up to several thousands of individuals in a single, small-scale PCR experiment (Zwaal *et al.*, 1993) as commonly applied in PCR-based *Petunia* insertion mutagenesis screens (Koes *et al.*, 1995; Vandenbussche *et al.*, 2003). The DNA samples prepared from the 3D pooled plant material each serve as a separate template for PCR amplification of the desired targets. A different amplification primer is used for every PCR reaction, carrying a unique four-nucleotide sample identification tag at the 5′ end (Keygene™ SeqTag; van Orsouw *et al.*, 2007). The PCR products are subsequently pooled, and unidirectionally sequenced using a modified GS20 protocol. The resulting sequences can be clustered according to identity and reduced to a single database line by 'reading' the various sequence identification tags encountered within each cluster. Querying the resulting database with the predefined profiles of individuals and families allows automatic assignment of sequences to their source.

For the *Petunia* experiment presented here, a classic 3D pooling system was used (Koes *et al.*, 1995; Vandenbussche *et al.*, 2003). The size of the analysed population, 1000



	X1	X2	X3	…	Y1	Y2	Y3	…	Z1	Z2	Z3	…
#A	1	0	0	0	1	0	0	0	1	0	0	0
famB	0	0	1	0	0	0	1	0	1	1	1	0

	X1	X2	X3	…	Y1	Y2	Y3	…	Z1	Z2	Z3	…
cluster1	1	0	0	0	1	0	0	0	1	0	0	0
cluster2	0	0	1	0	0	0	1	0	1	1	1	0
…	…	…	…	…	…	…	…	…	…	…	…	…

Figure 1. Multi-dimensional indexed massive parallel amplification and sequencing principle.
DNA templates are obtained by organizing the source material in a virtual 3D cubic set-up and pooling material from each dimension before DNA extraction. For example, all individuals in the *x*1 plane are pooled together, and this is performed similarly for all planes in all dimensions. Each individual DNA source in such a cube is therefore represented in three pools, once in every dimension, where the pool numbers reflect the spatial positioning of the DNA source in the 3D grid. Such 3D coordinate sets can also be translated to binary code as a series of ones and zeroes in a database query line, indicating the presence or absence of individual DNA templates in every pool of the library. Small groups of related individuals, such as families, can be profiled in the same way (e.g. famB). Templates are PCR-amplified from every DNA pool separately, using amplification primers containing pool number information encrypted as a four-nucleotide 'barcode' at its 5′ end. These amplification products are pooled together in one sample and sequenced by 454 sequencing. The resulting sequences are then clustered according to homology in the region between the amplification primers. Next, these clusters are then converted to single database lines containing a consensus sequence based on all sequences within each cluster, and the 5′ four-nucleotide sample identification tags are back-translated/converted to a series of numbers indicating the pool numbers encountered within each cluster. These database lines are then cross-searched against the originally defined profiles of individuals and known families within the population, leading to automatic annotation of the resulting clusters.

individuals, organized in a $10 \times 10 \times 10$ grid, is moderate, but nevertheless represents a highly complex sample due to the high copy number of the endogenous *Petunia dTph1* transposon system. Individual W138 plants, the standard line used in *Petunia* insertion mutagenesis screens, can contain up to an estimated 200 *dTph1* transposable element copies in either a homozygous or heterozygous state (Gerats *et al.*, 1990). Every generation, 5–10% of these transpose to a new location, creating a new insertion event, present in a single heterozygous individual (De Keukeleire *et al.*, 2001; Van den Broeck *et al.*, 1998). Upon selfing, such an insertion will segregate in the next generation. Thus, a population of 1000 plants could harbor 10 000–20 000 new insertions, hemizygously present in a single individual. In addition, as all W138 plants derive from a single plant (Doodeman *et al.*, 1984), ancestral insertions shared between all individuals are also encountered. A *dTph1* insertion library is usually composed of a number of small families (e.g. 100 families of 10 plants each), with every family originating from a self of an individual W138 plant. As such, unique insertions that occurred in the parental lines used to set up a library now segregate in their respective progenies. These are called family insertions, and are of great interest for the coupling of forward phenotypic screens with candidate gene insertion sequences.

Due to the complex genealogy of W138 insertion libraries, resulting in unique, family and ancestral insertion sites, considerable differences in copy number per insertion site exist within the population: the most extreme difference (1:2000) is between a new unique insertion event, which is hemizygously present in a single plant, and the immobile ancestral transposon insertion sites homozygous in all 1000 individuals. Family insertion copy numbers vary between these extremes, depending on the family size and the relatedness of the various families within the library. The mass amplification and sequencing of a complete *dThp1* insertion library is thus comparable with sequencing a non-normalized EST library, in which highly abundant, moderately expressed and minimally expressed ESTs correspond to ancestral, family and unique insertion events. As a result, the unique insertion flanking sequences are amplified much less efficiently than high copy number transposon flanking sequences. For reverse-genetics purposes, the high copy number class is of limited value, as it represents a small (and relatively immobile) set of insertion loci amplified from many to all individuals of the library. A minimum of around 50 such ancestral immobile loci were estimated to be present in previous populations (De Keukeleire *et al.*, 2001); therefore, this class of insertions unfortunately forms a large proportion of the total number of *dTph1* loci in a population ($50 \times 2 \times 1000 = 100\,000$). Therefore, the total number of sequences required to sufficiently resolve and annotate all low copy number flanking sequences (10 000–20 000) is very large, as omnipresent insertion sequences compete significantly for total sequencing capacity. Full coverage of all insertion flanking sequences in a limited number of sequence runs is therefore not expected. Moreover, a number of sequence clusters will remain unresolved, as one or more of the 3D coordinates will be missing. To at least partially reduce this problem, we implemented a normalization step for the amplified library before sequencing, in order to diminish the proportion of sequences derived from ancestral *dTph1* loci.

Results

3D indexed mass amplification and sequencing of Petunia dTph1 *transposon flanking sequences*

A graphical overview of the 3D indexed mass-amplification and sequence procedure, customized for analysis of *Petunia dTph1* transposon flanking sequences, is shown in Figure S1, and technical details are described in Experimental procedures. Briefly, we have applied a modified transposon display protocol (Van den Broeck *et al.*, 1998) to mass-amplify *dTph1* transposon flanking sequences from the 30 DNA pools harvested from the W138 library (Table S1). Four-nucleotide tags encrypting the 30 coordinates (Table S2) were incorporated in the transposon-derived primer used in the 2nd round of PCR amplification. The obtained PCR products were pooled for each dimension (*x*, *y* and *z*), resulting in three samples that were then subjected to one round of normalization using a hydroxyapatite column purification procedure (Bonaldo *et al.*, 1996). The selected single-stranded fraction, enriched for low copy number sequences, was made double-stranded by primer extension, and subsequently digested to enable GS20/454 adapter ligation. *Mse*I and *Mun*I sites were added to the transposon primer (containing the four-nucleotide sample identification tag) and the extension primer, respectively. The GS20/454 adapters A and B (Table S2) were modified to have sticky *Mse*I and *Mun*I ends instead of blunt ends, allowing unidirectional ligation of the sequencing annealing site in adapter A, immediately adjacent to the four-nucleotide tag in the transposon-specific primer. As a consequence, all amplified fragments originating from the same template have the same orientation, facilitating subsequent computational clustering. *Mse*I and *Mun*I were initially used for genomic DNA digestion, guaranteeing that no *Mse*I and *Mun*I sites will be present within the amplified transposon flanking sequences during digestion. The adapter-ligated mixture was then used as a template for a final PCR amplification step using amplification primers A and B (Table S2). A 4 μg aliquot of this mixture was then used as a template for two GS20 sequence runs, performed according to the manufacturer's instructions.

Clustering and automatic annotation of sequences to individual DNA sources

The combined output of the two GS20 runs yielded a total of 659 219 sequences, with an average read length of 101 bp. A BLAST search (Altschul *et al.*, 1990) was used for detection of the 5′ coded *Petunia dTph1* primer (NNNN-IR_{outw} primer, Table S2) at the start of the read; this could be recognized in 94.8% of the sequences. This result indicates that the large majority of the sequenced amplicons are derived from transposon flanking loci. The coordinates of the start of the BLAST hit were subsequently checked for the four-base sample identification tag. Using a PERL script, this tag could be automatically translated in one of the 30 possible pool coordinates in 96.4% of the sequences, and incorporated into a database line. Sequences with unrecognizable sample identification tags were marked as 'unknown' and removed at a later stage. Finally, primer and sample identification tag sequences were trimmed and all resulting sequences shorter than 20 bp were removed from the dataset. Such fragments originate from transposon insertion sites with an *Mse*I restriction site in close proximity to the *dTph1* insertion site and are of limited value for *in silico* screenings due to their short read length.

Clustering of the remaining dataset (611 000 sequences in total) resulted in the generation of 82 570 unique clusters, each with a variable number of homologous sequences. Each individual cluster was subsequently converted into a single database line containing a consensus sequence and all numerical data required for further classification and annotation. For each pool coordinate, the number of fragments encountered within the cluster carrying that coordinate was summed, and the total number of fragments in each cluster was counted. A distribution analysis based on these numbers showed that extreme differences in copy number exist between clusters, ranging from one to over 6000 copies (Figure 2a), as expected due to simultaneous amplification of unique and highly redundant insertion sites in a large population. Around 25% of the sequences (611 000 in total) belong to one of 113 clusters, with a copy number varying from 500 up to 6800. These clusters represent the so-called ancestral insertions that are present in the majority of the individuals within the population. These redundant insertion loci make up a considerable part of the total sequence capacity, but represent only a limited number of insertion loci. The largest number of clusters can be found in the lower copy number classes, corresponding to unique (both somatic and germinal) and recent family insertions, with the latter segregating in a limited number of related individuals within the library. For further analysis, the number of detected pools per pool axis (*x*, *y* and *z*) was calculated. With the latter information, sequence clusters were further classified and annotated according to 1D, 2D, 3D and mD (multiD) coordinate classes (Figure 2b). At this stage, 2759 of the clusters were discarded as they were composed solely of sequences previously marked as 'unknown' due to unreadable sequence tags. The 1D and 2D classes represent incomplete coordinate sets, missing a coordinate for two and one dimensions, respectively. The existence of incomplete coordinate sets is caused by either somatic insertion events or is due to under-sampling of the sequence library (see Discussion). On the other hand, 3D class clusters have exactly one *x*, one *y* and one *z* coordinate, and represent unique insertions present in one individual of the library. The mD class is composed of clusters having at least one *x*, *y* and *z* coordinate, excluding the 3D class clusters. These mD clusters therefore contain more than one pool coordinate in at least one of the dimensions, and correspond to insertion loci shared between at least two related individuals in the population.

Due to the 3D sampling strategy of the library, a sequence can only be assigned to an individual plant if it belongs to a cluster in which every dimension (*x*, *y* and *z*) is represented by at least one pool coordinate. All clusters fitting this criterion (3D and mD classes, 13 375 clusters in total) were therefore organized in a sub-database for final annotation.

In this set, a total of 5881 3D class clusters (Figure 2c), could be identified, which were automatically assigned a plant number between 1 and 1000 using the following formula:

$$((x-1)\times 10)+((y-1)\times 1)+((z-1)\times 100)+1,$$

with *x*, *y* and *z* all varying between 1 and 10 according to the library sampling used (Table S1).

The remaining set (mD class) represents so-called family insertions segregating in at least two related individuals within the library. As the relatedness of the lines in the library had been well documented during the previous two generations, a 'profile' for every family of plants in the library could be constructed as a translation of its spatial positioning within the library set-up (Table S1). Separate search profiles were created for small families and large families, for which the members display mainly first- and second-degree relatedness, respectively. These queries (see Experimental procedures) resulted in unambiguous annotation of around 4000 insertions as segregating in either one of the defined small or large families (Figure 2c, Tables S3 and S4). In addition, a small number of clusters did not contain sufficient information to be assigned to either of the two related search profiles. For these cases, a double annotation was applied, combining both possibilities in the sequence header. The remaining set of unassigned clusters was not further analysed, and presumably contains family insertions reflecting ancestral and undocumented relationships between individuals and families within the library. In a final step, only the annotated insertion flanking sequences from lines for which sufficient seeds are available for distribution were selected, and grouped together in a final BLAST-

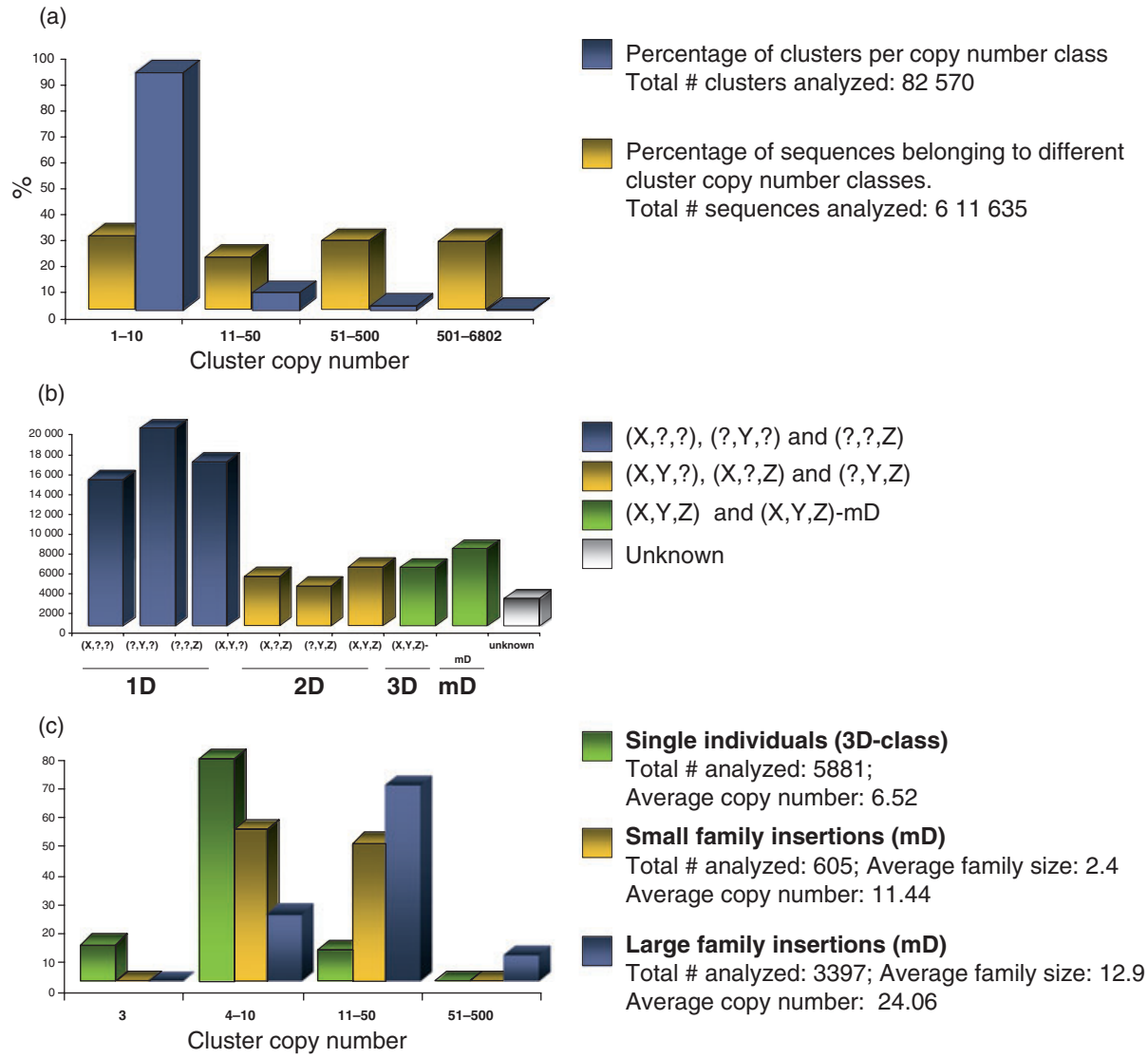


Figure 2. Distribution analysis of 3D indexed mass-amplified and sequenced *dTph1* transposon flanking sequence library.
(a) Cluster distribution analysis versus copy number classes.
(b) Number of clusters versus coordinate classes. Clusters have been classified according to the four possible dimensional classes. Question marks in the coordinates indicate the absence of pool coordinates for the respective dimension. The 'unknown' class indicates clusters containing only non-recognizable sequence tags.
(c) Number of assigned insertions versus cluster copy number classes. Total numbers of assigned insertions are indicated per category (total number analysed), together with the average copy number of clusters representing each type of insertion class.

searchable database, currently containing around 9500 non-redundant annotated transposon flanking sequences.

In silico *screens for insertions into genes of interest*

To evaluate whether the resulting annotated transposon flanking sequence database represents a reliable mutant identification source, we have performed an *in silico* screen for insertions into members of the *NAM*/*NAC* transcription factor family (Souer *et al.*, 1996). *NAC* genes have been identified as important regulators involved in various aspects of plant development (reviewed by Olsen *et al.*, 2005), and are easily recognized by their highly conserved NAC domain. The 22 known *Petunia NAC* family members were extracted from Genbank and subjected to a BLAST search against the insertion flanking sequence database. The search resulted in identification of nine putative insertions into eight *Petunia NAC* members (Figure 3a and Table S5). All these *dTph1* insertions reside in coding regions, which usually results in loss-of-function mutations. Five of these insertions correspond to known *Petunia NAC* members, while the remaining four represent insertions into

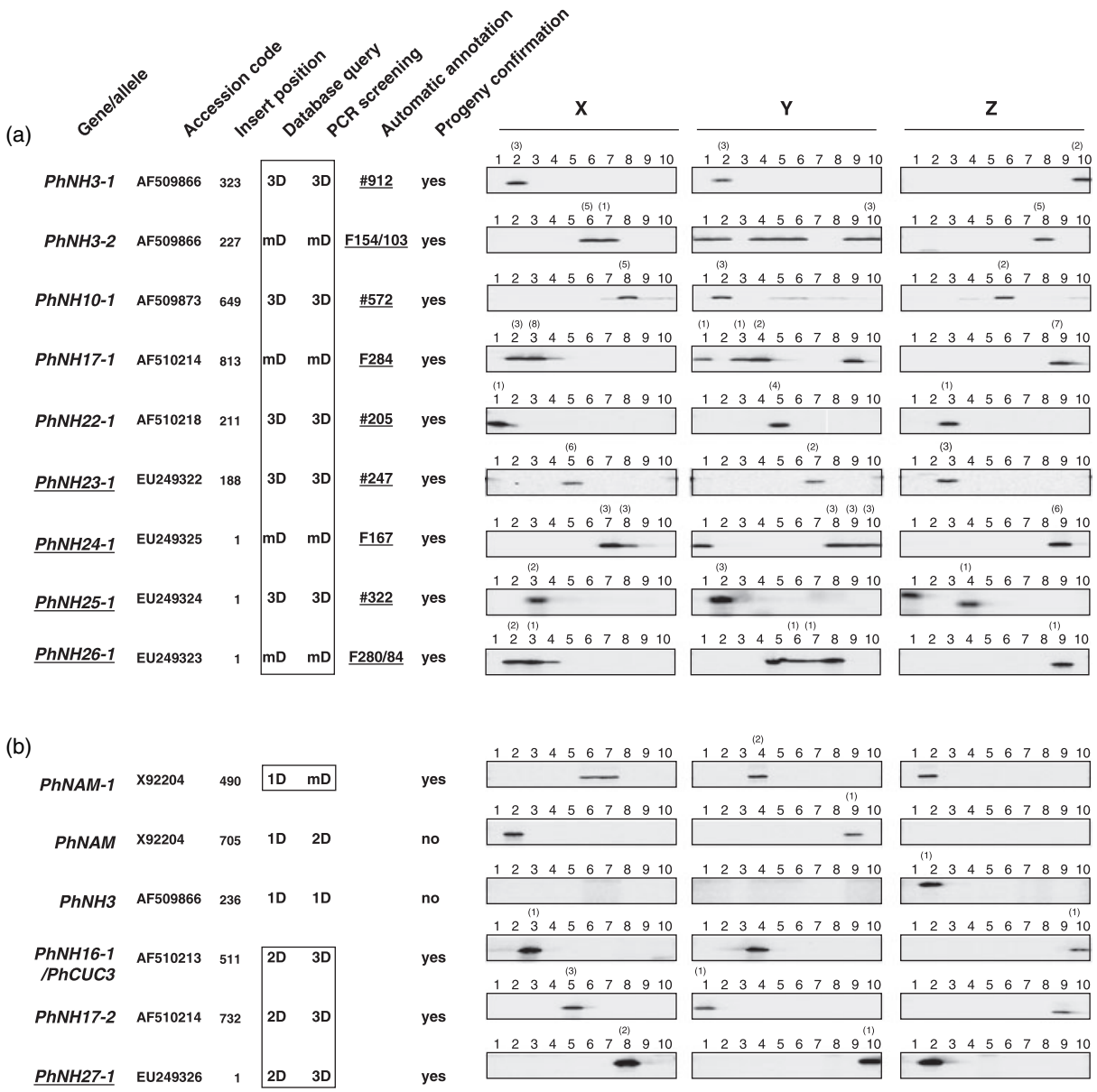


Figure 3. *In silico* and classic PCR screening for insertions into members of the *Petunia NAC* transcription factor family for complete gel images see Figure S2.
(a) Overview of all *in silico*-identified and automatically assigned insertions into *NAC* genes.
(b) Selection of the tested *in silico*-identified *NAC* insertions with incomplete (non-assignable) coordinate sets. The numbers above the gel images represent the 30 pool samples derived from the 3D pooling of 1000 individuals. Numbers in parentheses above the pool numbering reflect *in silico* screening results and show the copy number of transposon flanking sequences with corresponding pool numbers as encountered in the sequence collection. The results of the automatic annotation are shown as plant numbers for unique insertions, or as family names (F code) for segregating insertions. The gel images show the results of a conventional PCR screening as described previously (Vandenbussche *et al.*, 2003) using gene-specific primers designed to flank the predicted insertion sites. Insertion positions are relative to the start of corresponding Genbank accessions. 1D, 2D, 3D and mD indicate the coordinate sets after database screening (*in silico*) and classic PCR screening. Assignable coordinate sets have been boxed. New NAC members isolated during this screening are underlined.

new *NAC* family members (named PhNH23–26), based on sequence conservation in their NAM domain (Figure S3), illustrating additional benefits of the database as a gene discovery source. The coordinates of five of the nine putative insertions identify a unique single plant (3D class insertion) within the population, while the remaining four insertions represent so-called family insertions (mD class), segregating in a small group of genetically related plants. In addition, three putative insertion events were identified that had one of their coordinates missing (2D hits), while 15 putative insertions were identified as 1D sets (not shown). Incomplete coordinate sets may represent somatic insertion events, present in only one or two of the three samples taken from each individual, or might represent true germinal insertion events for which one or more of the coordinates is missing for technical reasons. To investigate this further, we

included three 2D and three 1D sets in the analysis presented below.

As an independent confirmation of the automatic assignment of insertions to individuals and small families, we screened the 30 DNA pools of the library using a conventional gene-specific screening method (Vandenbussche *et al.*, 2003) for each of the nine automatically assigned insertions (Figure 3a) and six selected incomplete sets (Figure 3b) (see also Figure S2a–e). All *in silico*-assigned insertion events could be confirmed by PCR screening, as shown by the appearance of specific amplification products in all of the identified pool coordinates. Further, presumably due to the higher sensitivity of the gene-specific screening method, the missing coordinate for all three 2D sets could be determined (identifying a new gene: PhNH27-1; Figure 3b), but this was not possible for two of the three 1D hits (Figure 3b). These results confirm both the occurrence of somatic insertion events, and the incompleteness of clusters due to under-sampling of the sequence library. Similarly, for three family insertions, additional positive pools were identified that were not found in the sequence database; these pools fitted within the previously identified family profile. These results clearly indicate that not all templates are amplified or not all amplified templates are represented in our current sequence collection. Finally, we were able to confirm the presence of the respective insertion events in progenies of all positive individuals or families identified in this screen (results not shown); a number of these lines are currently being analysed for phenotypes.

Discussion

The use of 5′ coded primers has been proposed previously to distinguish a limited number of DNA sources within a 454 sequence template (Binladen *et al.*, 2007; van Orsouw *et al.*, 2007). Here we show that the combined use of 5′ coded primers with a multidimensional pooling and annotation strategy results in a extremely powerful method, enabling highly complex templates to be analysed by 454 sequencing while the source is still identifiable, and with a minimal labour investment in terms of template preparation and processing. This approach can be applied to any system for which multiple templates from multiple sources have to be analysed on a truly large scale. The generation of hundreds of thousands of sequences per 454 run provides sufficient over-sampling capacity to analyse and annotate tens of thousands of templates simultaneously.

We developed and applied this procedure to generate a database of annotated transposon flanking sequences, mass-amplified from a *Petunia* W138 library of 1000 individuals, which can then be used for *in silico* reverse-genetics screens. However, the nature of the endogenous *dTph1* transposon system in *Petunia* created some extra complicating factors.

The extreme differences in copy number between ancestral and new unique insertion sites within such a library made it necessary to normalize the samples before sequencing. Even so, 25% of the final sequence reads appeared to be derived from insertions present in many to all individuals of the library. Other normalization procedures might turn out to be more efficient (see, for example, Cheung *et al.*, 2006; Shagin *et al.*, 2002). Other undesired sequence tags are those that are derived from somatic insertions and thus will not be transmitted to the next generation. The presence of these sequences in the final database would lead to false identification of putative insertion lines. In *Petunia*, it has been demonstrated that a 3D pooling strategy allows discrimination between somatic and germinal insertion events (Koes *et al.*, 1995). Unfortunately, a 3D indexed pooling strategy implies that a sequence can only be annotated if it has been amplified and sequenced at least three times independently in the total sequence collection, at least once from every dimensional pool.

It is obvious that, with the same total sequence capacity, a 2D pooling strategy would have resulted in a larger number of annotated sequences, as two coordinates each time would have sufficed to appoint a sequence to a source. We nevertheless chose a 3D indexed mass amplification and sequencing strategy, so as to be able to exclude all somatically derived transposon flanking sequences from the database. This problem is specific to *Petunia*; for most other applications, a 2D pooling strategy might be a better choice.

Despite the complications inherent in *dTph1* transposon biology, we identified almost 10 000 different insertion flanking sequences that could be automatically assigned to specific individuals of the library using only 30 DNA preparations, 70 PCR reactions and two GS20 sequencing runs. Almost 6000 (5881) of the sequence tags represent unique insertions present in only one individual of the library. Using defined family history profiles based on three generations of known family relationships between the various individuals, we could automatically annotate a further 4000 insertions as segregating in small families in the library.

The successful assignment of these family insertions in this study also indicates that the methodology is well suited to large-scale parallel mapping of family-specific genotypes in heterogeneous populations. Combined with the fast progress in 454 sequencing technology in terms of both capacity and read length, and with further improvement and application of the proposed strategy, we expect that creation of a saturated insertion sequence library is realistic for *Petunia* as well as for other model systems. As a starting point, we have organized the approximately 10 000 identified insertion sequence tags in a database. BLAST searches for insertions into genes of interest will be performed on request, and seeds of positively identified insertion lines will

be provided free for academic purposes. We expect this to be a valuable resource for functional genomics research in *Petunia* and other Solanaceae.

Experimental procedures

Library set-up

Seeds representing 1000 progenies from self-pollinations (mainly W138 but also W138/Mitchell hybrids) were sown, and usually one individual of each progeny was grown for the transposon library. These 1000 plants were grown under standard greenhouse conditions in trays, and after harvesting for DNA sampling were transferred to individual pots and maintained until sufficient seeds were produced, resulting from manual pollinations. A schematic representation of the library set-up is shown in Table S1. Equal colouring within blocks (*z* dimension) between individuals indicates a family relationship in the 2nd degree. Identical parental accession codes at different positions indicate direct sibling relationships.

DNA preparation

The 1000 selected library plants were sampled in a 3D fashion, resulting in 30 DNA samples representing 100 plants each, as described previously (Koes *et al.*, 1995; Vandenbussche *et al.*, 2003). The DNA samples were further treated using a modified version of the transposon display method originally developed to selectively amplify all *dTph1* flanking sequences from various samples of a one single plant (Van den Broeck *et al.*, 1998).

A 5 μg DNA aliquot for each pool sample was incubated with *Mun*I (Biolabs; http://www.neb.com/nebecomm) and *Mse*I (Biolabs; http://www.neb.com/nebecomm) for 1.5 h at 37°C in 70 μl restriction mix (5 units *Mun*I, 5 units *Mse*I, 1× NEB4 and 1× BSA). Adapters were ligated by incubation for 3 h at 37°C after addition of 30 μl ligation mix (15 Weiss units of T4 DNA ligase (Fermentas; http://www.fermentas.com), 40 pmol biotinylated *Mun*I adapter, 400 pmol *Mse*I adapter, 1 mM ATP, 1× NEB4 and 1× BSA).

Selection of biotinylated DNA fragments to enrich for transposon flanking fragments

To remove the excess of biotinylated adapters, DNA samples were purified using the QIAquick PCR purification kit (Qiagen, http://www.qiagen.com/) and DNA was eluted using 50 μl EB buffer (supplied elution buffer; Qiagen PCR purification kit). DNA samples were incubated with approximately 0.1 mg MyOne streptavidine beads C1 (Dynal; http://www.invitrogen.com/dynal) for 1 h on a rotator at room temperature, in 500 μl binding buffer (10 mM Tris–Cl pH 8, 2 M NaCl, 1 mM EDTA and 0.1% Triton X-100). Beads were collected with a magnet and the supernatant removed. Beads were washed once using 200 μl STEX buffer (10 mM Tris–Cl pH 8, 1 M NaCl, 1 mM EDTA and 0.1% Triton X-100), and transferred to another tube prior to three additional washes with 200 μl STEX buffer each. Finally the beads were resuspended in 50 μl $T_{01}E$ buffer (10 mM Tris–Cl pH 8, 0.1 mM EDTA) and transferred to another tube.

Selective pre-amplification

To enrich for transposon-flanking sequences, DNA–bead samples were subjected to two rounds of PCR amplification. For the first PCR amplification, 2 μl of DNA–bead sample was mixed with 6 pmol *Mun*I-ACAC- and 6 pmol *Mse*I primer (Table S2) in a PCR buffer with 0.6 units of Red Hot DNA polymerase (Abgene; http://www.abgene.com), 7.5 mM Tris–Cl pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 2.5 mM $MgCl_2$ and 0.2 mM dNTPs in a final volume of 20 μl. The samples were incubated in a PE9600 (Perkin Elmer; http://www.appliedbiosystems.com) according to the following PCR profile: one cycle comprising a 15 sec step at 94°C, a 30 sec step at 65°C and a 60 sec step at 72°C; the annealing temperature was then lowered each cycle by 0.7°C for 13 cycles, and then kept at 56°C for another 22 cycles. The resulting PCR products were checked by electrophoresis of 5 μl of reaction mix on a 1.5% agarose gel (a weak low-molecular-weight smear should be visible). The remaining 15 μl was diluted 10 times with H_2O.

Specific PCR amplification of dTph1 *flanking sequences*

To amplify the *dTph1* flanking sequences and to incorporate the pool-specific four-nucleotide sample identification tag, 5 μl of the 10-fold diluted pre-amplified material was mixed with 15 pmol NNNN-IR_{outw} primer and 15 pmol Mse primer (Table S2) in a PCR buffer with 1 unit Red Hot DNA polymerase (Abgene), 7.5 mM Tris–Cl pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 2.5 mM $MgCl_2$ and 0.2 mM dNTPs in a final volume of 50 μl. The samples were incubated in a PE9600 (Perkin Elmer) according to the following PCR profile: one cycle comprising a 15 sec step at 94°C, a 30 sec step at 65°C and a 60 sec step at 72°C; the annealing temperature was then lowered each cycle by 0.7°C for 13 cycles, and then kept at 56°C for another 22 cycles.

The resulting PCR products from the ten samples from each dimension were pooled to create three samples: an *x*, *y* and *z* sample.

Normalization

In order to optimize the amount of unique fragments against a backdrop of fragments shared by many to all individuals, we normalized the three pooled samples. The procedure involves hybridization and purification steps to obtain single-stranded DNA. The approach relies on the differential re-annealing kinetics between highly abundant and rare DNA templates, by which the single-stranded fraction will become enriched for low copy number templates. The single-stranded fraction was isolated by hydroxyapatite chromatography (Bonaldo *et al.*, 1996).

The pooled PCR products (*x*, *y* and *z* samples, containing around 10 μg of DNA each) were precipitated and dissolved in 30 μl formamide. The DNA was melted at 80°C for 3 min (under mineral oil) and hybridized for 16 h at 30°C, as described in Bonaldo *et al.* (1996), in a final volume of 50 μl. Thereafter, the samples were diluted by addition of 500 μl pre-warmed (60°C) loading buffer (0.04 M sodium phosphate buffer, pH 6.8), and single-stranded DNA was purified by hydroxyapatite chromatography at 60°C, using a 0.1 ml column pre-equilibrated with the loading buffer. After a 2 ml wash with loading buffer, the single-stranded DNA was eluted from the column using eight repeats of 100 μl pre-warmed 0.12 M sodium phosphate buffer (pH 6.8), and elution fractions 2–8 were pooled. The phosphate buffer was removed using nucleospin columns (Macherey Nagel; http://www.mn-net.com). The DNA was eluted four times using 200 μl NE buffer, precipitated with iso-propanol, and finally dissolved in 50 μl H_2O for each elution. The effectiveness of the normalization was verified by a PCR competition experiment between a high copy number flanking sequence and a low copy sequence, and a

transposon display experiment on normalized and non-normalized samples (results not shown). In view of the results, further optimization of the normalization would be useful, and could be achieved by varying parameters or using other procedures (see, for example, Cheung *et al.*, 2006; Shagin *et al.*, 2002).

GS20 template preparation and sequencing

Primer extension. For conversion to double-stranded DNA, 50 μl of single-stranded DNA was mixed with 50 pmol extension primer (Table S2) in a PCR buffer with 1 unit Platinum Taq DNA polymerase (Invitrogen, http://www.invitrogen.com/), 20 mM Tris–Cl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.2 mM dNTPs in a final volume of 75 μl. The three samples were then incubated in a PE9600 (Perkin Elmer) for 2 min at 94°C, 1 min at 56°C and 10 min at 72°C to complete the primer extension.

Digestion and GS20 adapter ligation. The three normalized (double-stranded) DNA pools were digested with *Mun*I and *Mse*I to enable directional GS20 adapter ligation. The total amount of normalized double-stranded DNA was incubated for 1.5 h at 37°C in 90 μl restriction mix (20 units *Mun*I, 20 units *Mse*I, 1× NEB4 and 1× BSA). The GS20 adapters were ligated by incubation for 4 h at 37°C after addition of 20 μl ligation mix (15 Weiss units of T4 DNA ligase, 200 pmol GS20-*Mun*I adapter B, 200 pmol GS20-*Mse*I adapter A, 1 mM ATP, 1× NEB4 and 1× BSA).

If there is enough DNA (a few micrograms; when normalization isn't required), it is possible to ligate the BioTEG-GS20-*Mun*I adapter B (Table S2) instead of the non-biotinylated adapter. In this case, the samples can be pooled after adapter ligation to create one 'superpool', and are ready for the 'library immobilization step' of the GS20 sequencing protocol.

PCR re-amplification before sequencing. After normalization, we re-amplified the ligation mixture to obtain sufficient and biotinylated DNA for GS20 sequencing. Five microlitres of template were mixed with 15 pmol of adapter primer A and 15 pmol BioTEG-adapter B primer (Table S2; http://www.biolegio.com) in a PCR buffer with 1 unit of Red Hot DNA polymerase (Abgene), 7.5 mM Tris–Cl pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 2 mM $MgCl_2$ and 0.2 mM dNTPs in a final volume of 50 μl. The samples were incubated in a PE9600 (Perkin Elmer) according to the following PCR profile: one cycle with a 15 sec step at 94°C, a 30 sec step at 65°C and a 60 sec step at 72°C; the annealing temperature was then lowered each cycle by 0.7°C for 13 cycles, and then kept at 56°C for another 22 cycles.

The samples were pooled to create one superpool. Approximately 4 μg of the superpool fragments were used as input for GS20 library construction. Several adaptations to the standard library preparation protocol were made. First, no shearing, end-polishing, phosphorylation or ligation of the A and B adapters was carried out; the samples entered at the 'library immobilization step' of the protocol. Second, the Bst DNA polymerase fill-in step of the protocol was omitted.

Sequencing. After library construction, a titration run was performed to estimate the optimal amount of single-stranded fragments containing the A and the B adapter for the actual sequencing run. After the titration run, emulsion PCR and bead enrichment were carried out according to the standard GS20 protocol (Roche Applied Science; http://www.roche-applied-science.com). Two full picotitre plates (70 × 75 mm) with two regions each were used. Sequencing was performed according to the manufacturer's instructions (Roche Applied Science).

Clustering of the GS20 sequence output

The cleaned and trimmed dataset was clustered using a combined approach. The first step was a fast clustering using CDHIT (Weizhong *et al.*, 2001), using a percentage identity of 93%. From each resulting cluster, a consensus fragment was generated as a representative. All CDHIT cluster representatives were further iteratively clustered using BLASTCLUST (Altschul *et al.*, 1990) (percentage identity of 93%), until the number of clusters did not further decrease.

Annotation of segregating insertions in the library

For each of the 30 pool coordinates, these small and large family profiles (Table S3 and S4) contain the value *x* (where the number of fragments within the cluster assigned to the given pool can be 0 or more), 0 (where the number of fragments within the cluster assigned to the given pool is 0), >1 (which is the number of fragments within the cluster assigned to the given pool) or *A* (where the sum of the number of fragments encountered within the pools with this value *A* is 1).

A number of search profiles, especially in the large families, display closely related patterns, due to inter-mixed positional arrangement of the corresponding families within the library. For a more unambiguous distinction between these patterns, we included the operators '*A*' and '>1' in the search profiles of some families.

The search profiles were translated into SQL queries using a PERL script. The script executed the queries, and all resulting entries were annotated as a family insertion (small or large class). This step was consecutively performed for small and large families, and only the remaining non-annotated insertions were searched for in the 2nd round. This strict order of extraction is necessary as the majority of small family profiles reside within larger family profiles.

Sequence deposition

Sequence data for *PhNH23*, *PhNH24*, *PhNH25*, *PhNH26* and *PhNH27* have been deposited with the EMBL/Genbank libraries under accession numbers EU249322–EU249326.

Patents and trademarks

The method for high-throughput screening of transposon tagging populations using massively parallel sequence identification of insertion sites and KeyGene™ SeqTag technology are covered by patents and patent applications owned by Keygene NV (http://www.keygene.com). Application for trademark registration for Seq Tag technology has been filed by Keygene NV.

Acknowledgements

We thank Hans Sommer and Zsuzsanna Schwarz-Sommer (both Max Planck Institute, Cologne, Germany) for sharing their normalization protocol. M.V. was funded by a HORIZON grant (050-71-036) from the Netherlands Genomics Initiative. A.S.R. was funded by the Netherlands Organization for Scientific Research (grant 814.02.009).

Supplementary Material

The following supplementary material is available for this article online:

Figure S1. Overview of 3D indexed mass amplification and directional sequencing of *Petunia* right border *dTph1* flanking fragments.

Figure S2. PCR confirmation of *in silico*-predicted *NAC* insertion alleles for 3D pooled DNA samples from the library.

Figure S3. Protein alignment of translated *dTph1* flanking sequences representing new *Petunia NAC* members, together with a selection of known *Petunia NAC* genes.

Table S1. Library set-up.

Table S2. Oligo and adapter sequences.

Table S3. Small family search profiles.

Table S4. Large family search profiles.

Table S5. Database representation of the nine *in silico*-identified *NAC* insertions.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

Alonso, J.M., Stepanova, A.N., Leisse, T.J. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Binladen, J., Gilbert, M.T., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. and Willerslev, E. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS*, **14**, e197.

Bonaldo, M., Lennon, G. and Bento Soares, M. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* **6**, 791–806.

Cheung, F., Haas, B.J., Goldberg, S.M.D., May, G.D., Xiao, Y. and Town, C.D. (2006) Sequencing *Medicago truncatula* expressed sequenced tags using 454 life sciences technology. *BMC Genomics*, **7**, 272–282.

De Keukeleire, P., Maes, T., Sauer, M., Zethof, J., Van Montagu, M. and Gerats, T. (2001) Analysis by transposon display of the behavior of the dTph1 element family during ontogeny and inbreeding of *Petunia hybrida*. *Mol. Genet. Genomics*, **265**, 72–81.

Doodeman, M., Boersma, E.A., Koomen, W. and Bianchi, F. (1984) Genetic analysis of instability in *Petunia hybrida*: a highly unstable mutation induced by a transposable element in the An1 locus for flower colour. *Theor. Appl. Genet.* **67**, 345–355.

Droc, G., Ruiz, M., Larmande, P., Pereira, A., Piffanelli, P., Morel, J.B., Dievart, A., Courtois, B., Guiderdoni, E. and Perin, C. (2006) OryGenesDB: a database for rice reverse genetics. *Nucleic Acids Res.* **34**, D736–D740.

van Enckevort, L.J., Droc, G., Piffanelli, P. *et al.* (2005) EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol. Biol.* **59**, 99–110.

Gerats, T. and Vandenbussche, M. (2005) A model for comparative research: Petunia. *Trends Plant Sci.* **10**, 251–256.

Gerats, A.G., Huits, H., Vrijlandt, E., Marana, C., Souer, E. and Beld, M. (1990) Molecular characterization of a nonautonomous transposable element (dTph1) of petunia. *Plant Cell*, **2**, 1121–1128.

Koes, R., Souer, E., Houwelingen, A. *et al.* (1995) Targeted gene inactivation in Petunia by PCR-based selection of transposon insertion mutants. *Proc. Natl Acad. Sci. USA*, **92**, 8149–8153.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* **37**, 301–314.

Margulies, M., Egholm, M., Altman, W.E. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Olsen, A.N., Ernst, H.A., Leggio, L.L. and Skriver, K. (2005) NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci.* **10**, 79–87.

van Orsouw, N.J., Hogers, R.C.J., Janssen, A. *et al.* (2007) Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE*, **2**(11), e1172.

Settles, A.M., Holding, D.R., Tan, B.C. *et al.* (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics*, **8**, 116.

Shagin, D.A., Rebrikov, D.V., Kozhemyako, V.B., Altshuler, I.M., Shcheglov, A.S., Zhulidov, P.A., Bogdanova, E.A., Staroverov, D.B., Rasskazov, V.A. and Lukyanov, S. (2002) A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Res.* **12**, 1935–1942.

Souer, E., van Houwelingen, A., Kloos, D., Mol, J. and Koes, R. (1996) The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*, **85**, 159–170.

Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De Keukeleire, P., D'Hauw, M., Van Montagu, M. and Gerats, T. (1998) Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* **13**, 121–129.

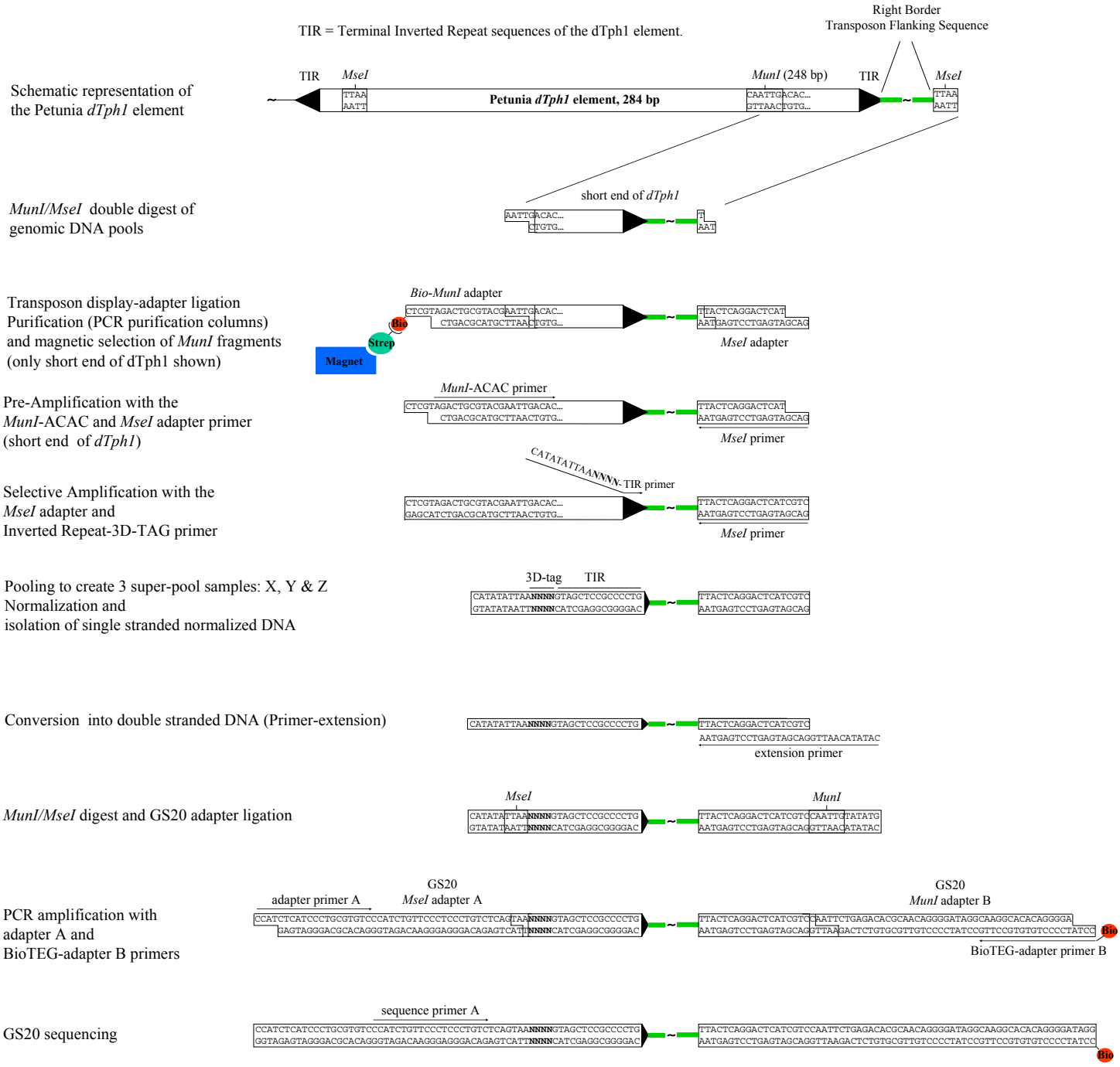
Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornielli, G.B., Pezzotti, M., Ferrario, S., Angenent, G.C. and Gerats, T. (2003) Toward the analysis of the Petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require SEPALLATA-like MADS box genes in Petunia. *Plant Cell*, **15**, 2680–2693.

Weizhong, L., Jaroszewski, L. and Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein database. *Bioinformatics*, **17**, 282–283.

Yazaki, J., Kojima, K., Suzuki, K., Kishimoto, N. and Kikuchi, S. (2004) The rice PIPELINE: a unification tool for plant functional genomics. *Nucleic Acids Res.* **32**, D383–D387.

Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T. and Plasterk, R.H. (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl Acad. Sci. USA*, **90**, 7431–7435.

The EMBL/Genbank accession numbers for the *PhNH23–27* sequences are EU249322–EU249326, respectively.



Supplementary Figure1: **Overview of 3D-indexed mass amplification and directional sequencing of petunia right border *dtph1* flanking fragments.**

Supplementary Figures 2A-E: PCR confirmation of nam insertion alleles on 3D pooled DNA samples of the library, complete gel images.

M = Size marker
Arrows indicate fragments representing NAC insertion alleles. Sequences of the gene specific primers used in these screenings are provided in Supplementary Table 2.
Corresponding insertion alleles are indicated.

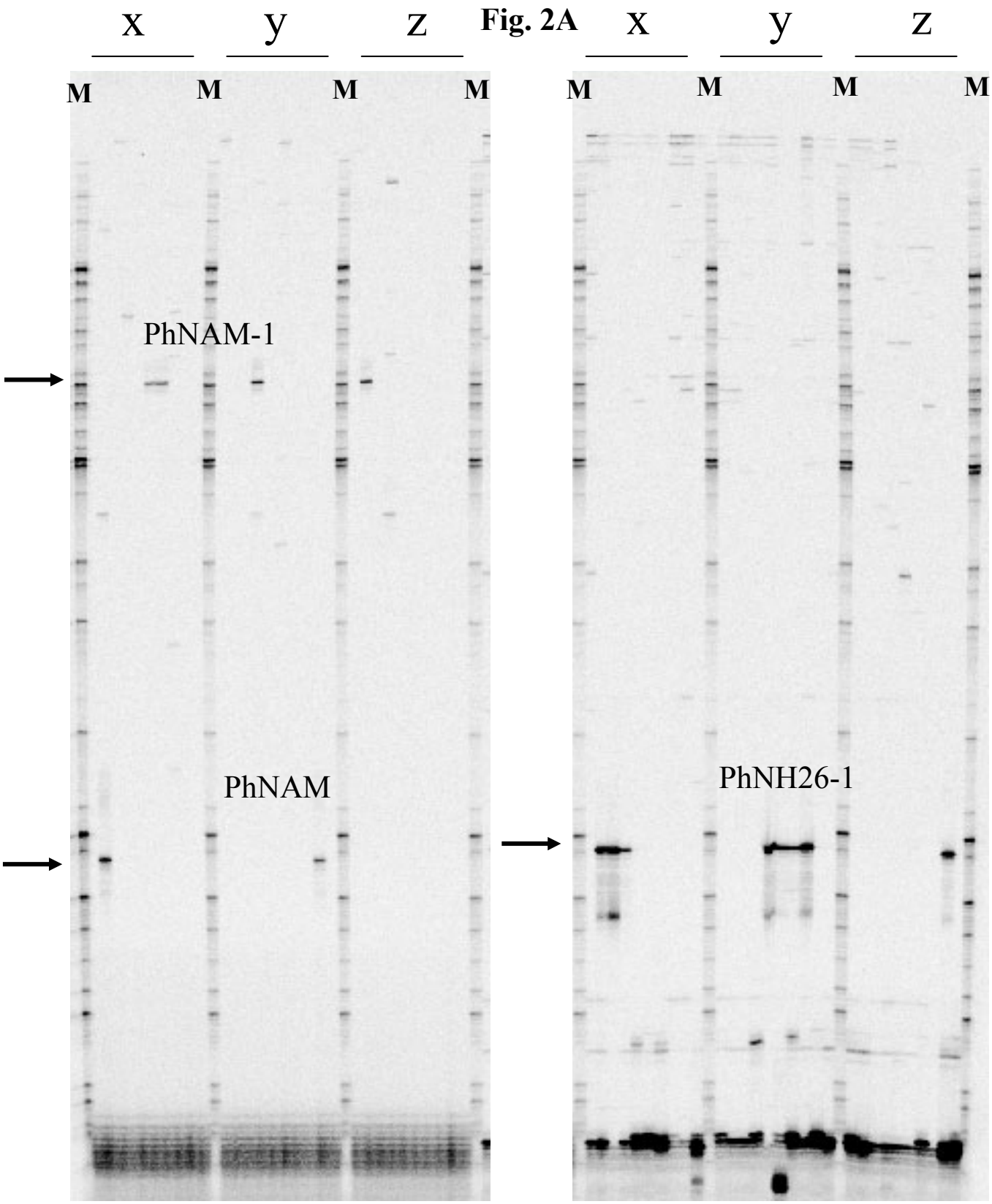


Fig. 2A

Fig. 2B

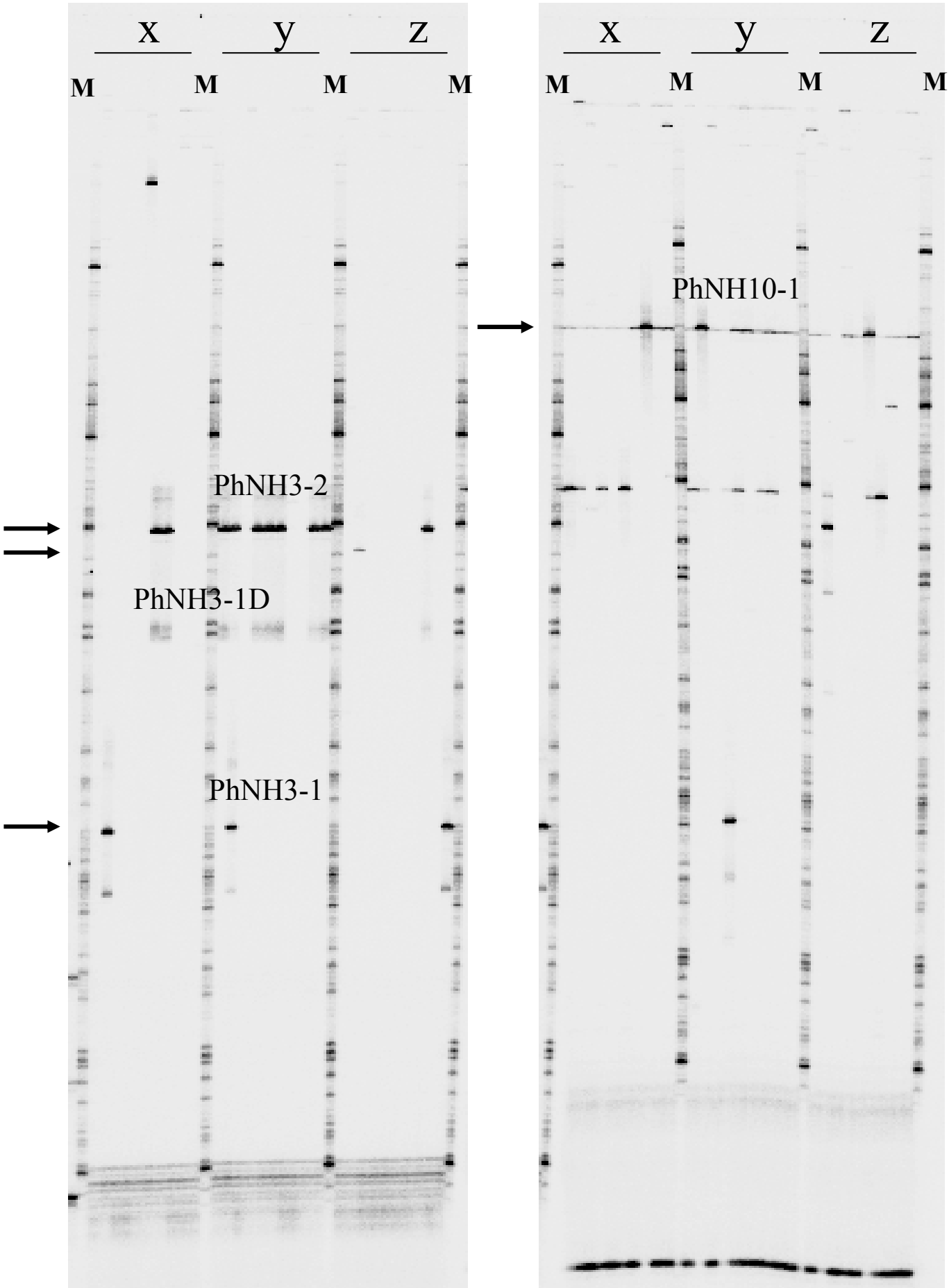


Fig. 2C

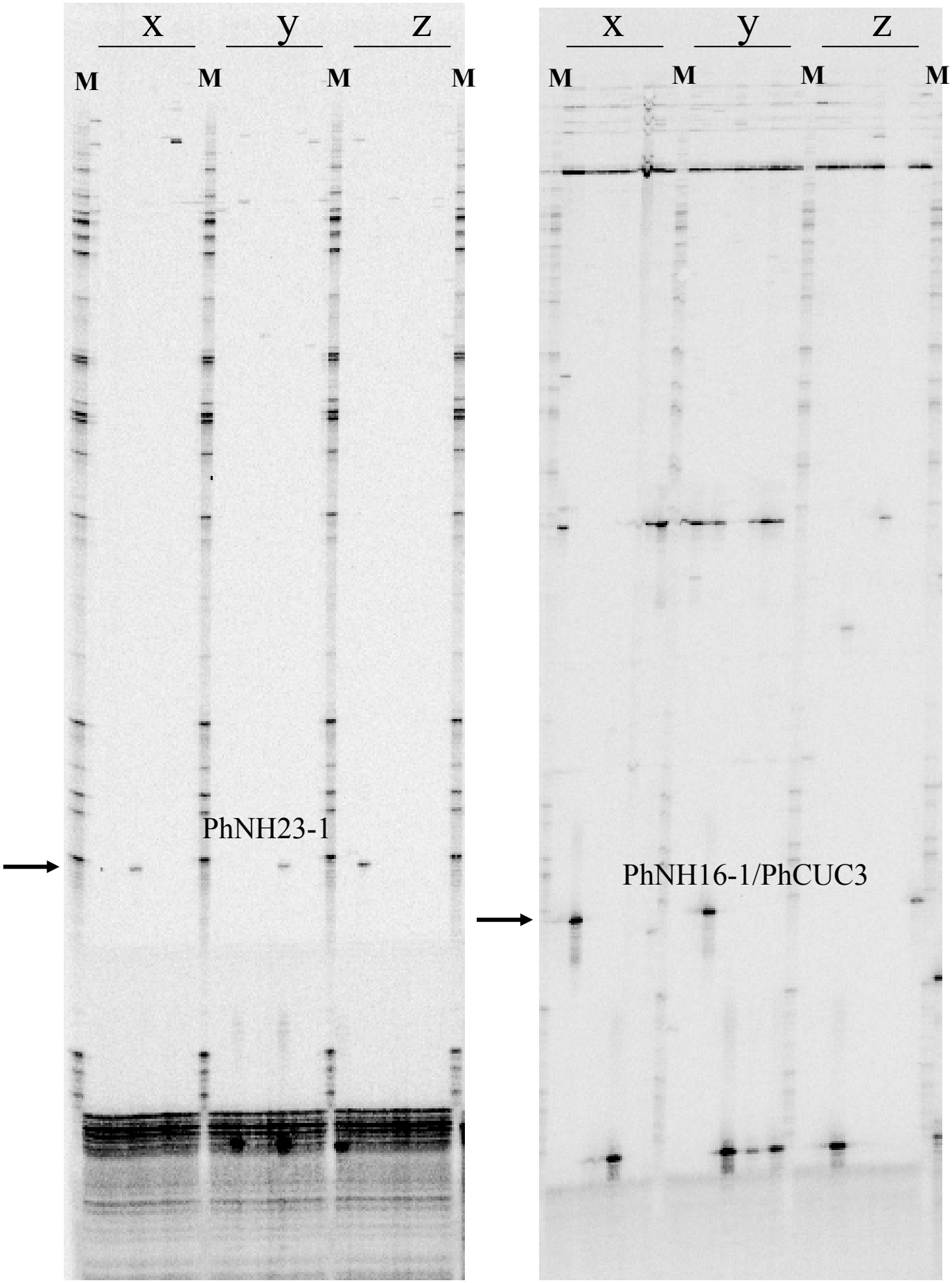


Fig. 2D

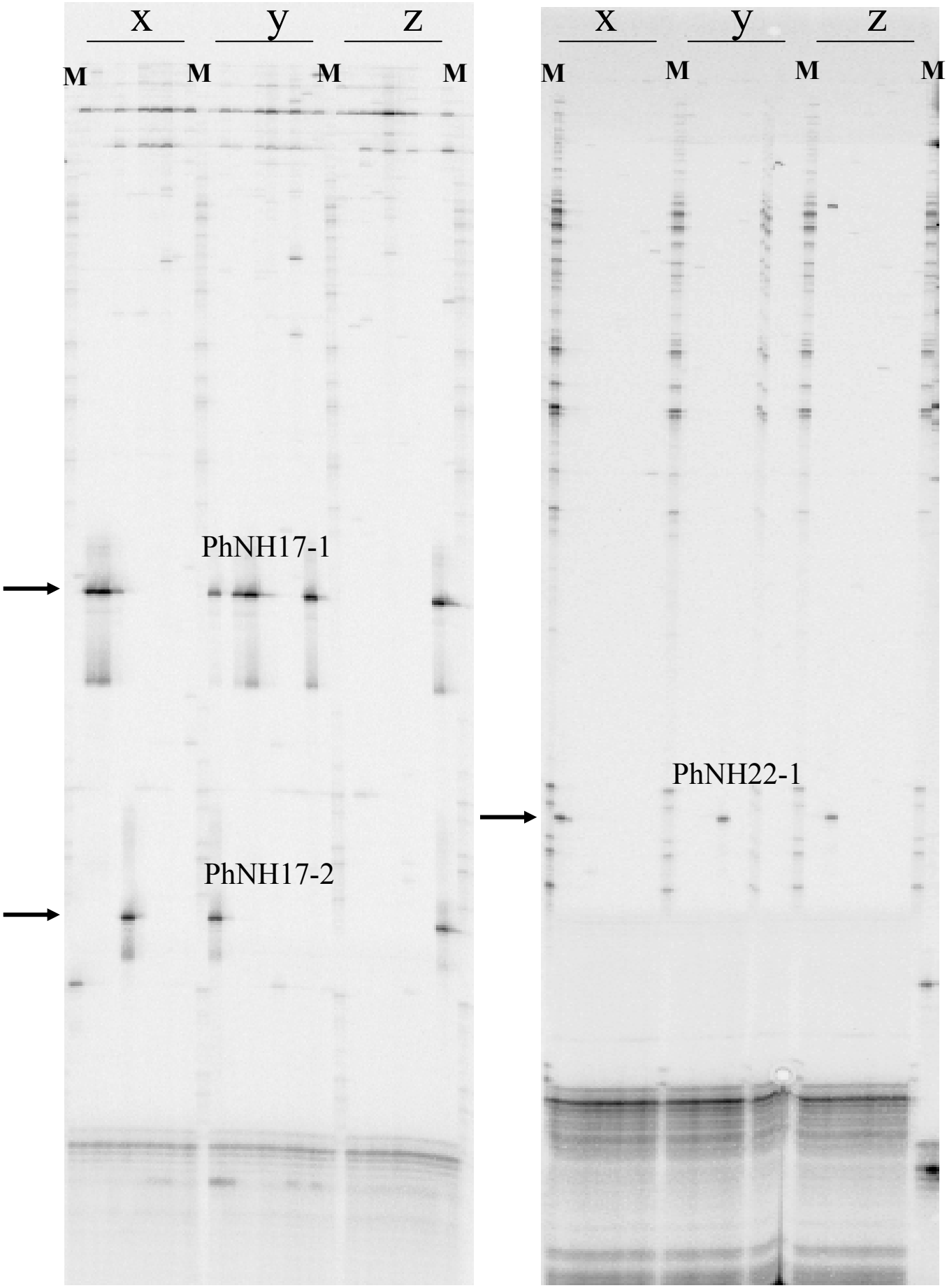
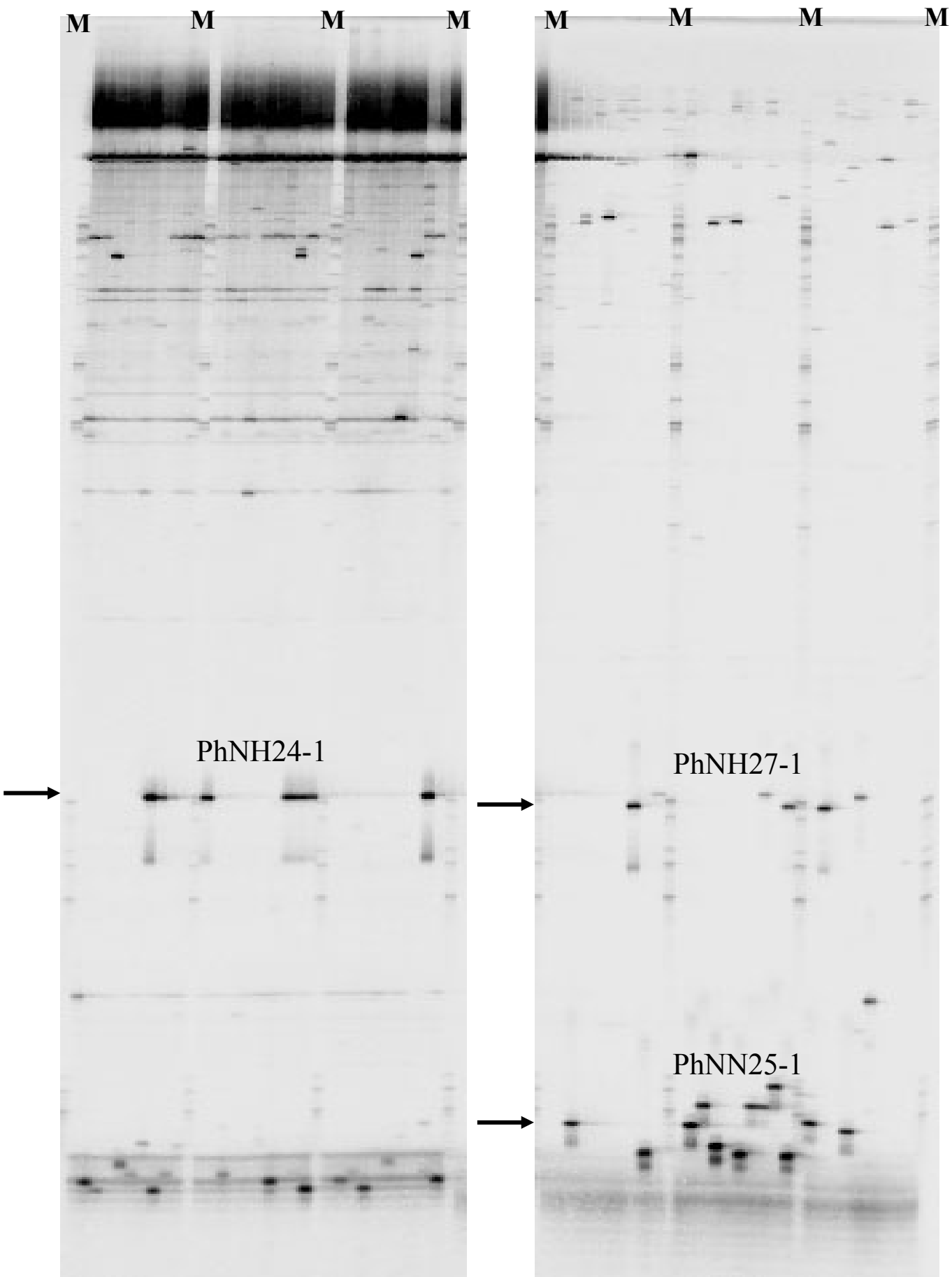
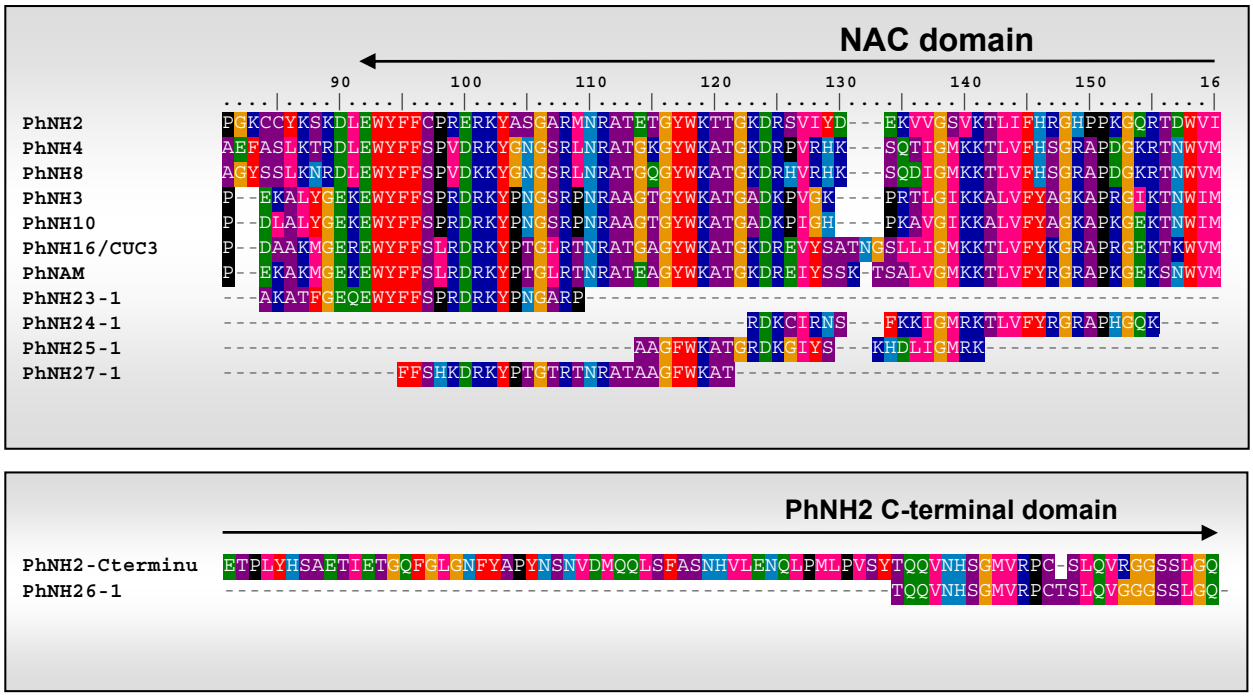


Fig. 2E





Supplementary Figure 3: Protein alignment of translated *dTph1* flanking sequences representing new *Petunia* NAC members together with a selection of known Petunia NAC genes.

Supplementary Table 1: Library set-up

	X1		X2		X3		X4		X5	
	Z1									
Y1	229-7(12B)	1	K262-9-W	11	K270-6-W	21	K111-6-W(12B)	31	K262-9-W	41
Y2	229-8	2	229-10	12	K272-7-W	22	K111-5-W	32	K262-3-W	42
Y3	229-9-W	3	229-23-W	13	K234:14	23	K111-3-W	33	K262-2	43
Y4	229-10	4	229-24-W	14	K234:15	24	K111-1-W	34	K262-1	44
Y5	229-11-W	5	212-10-W	15	K234:16	25	216-6 (12D)	35	E210-9 (14D)	45
Y6	229-16	6	113-4-W	16	113-7-W	26	212-5 (12D)	36	198-15	46
Y7	229-16	7	212-18 (13A)	17	113-4-W	27	198-15	37	E210-5	47
Y8	229-17	8	K218-8(12C)	18	113-3-W	28	K218-9(12C)	38	K218-7	48
Y9	113-2-W	9	K218-6	19	113-2-W	29	K218-7	39	198-15	49
Y10	K218-6	10	[illegible]	20	K111-7-W	30	K218-8	40	[illegible]	50

	X6		X7		X8		X9		X10		
	Z1										
Y1	229-7(12B)	51	229-21	61	K270-6-W	71	K111-6-W(12B)	81	K262-9-W	91	Y1
Y2	229-8	52	K272-7-W	62	K272-7-W	72	K111-5-W	82	K262-3-W	92	Y2
Y3	229-9-W	53	229-23-W	63	K111-3-W	73	K111-3-W	83	K262-2	93	Y3
Y4	229-10	54	229-24-W	64	K111-1-W	74	K111-1-W	84	K262-1	94	Y4
Y5	229-11-W	55	212-10-W	65	212-10-W	75	216-6 (12D)	85	E210-9 (14D)	95	Y5
Y6	229-11-W	56	212-10-W	66	113-7-W	76	212-5 (12D)	86	E210-6	96	Y6
Y7	229-16	57	212-18 (13a)	67	113-4-W	77	198-15	87	E210-5	97	Y7
Y8	229-17	58	113-3-W	68	113-3-W	78	K218-9(12C)	88	E210-2	98	Y8
Y9	229-18	59	113-2-W	69	113-2-W	79	113-2-W	89	[illegible]	99	Y9
Y10	K111-7-W	60	K111-7-W	70	K111-7-W	80	K218-6	90	[illegible]	100	Y10

	X1		X2		X3		X4		X5	
	Z2									
Y1	116-14X120-8	101	120-23	111	116-14X120-8	121	318-16X198-7	131	318-16X198-7	141
Y2	116-23X199-8	102	116-1X199-9	112	116-15X120-8	122	318-20	132	318-17-W	142
Y3	116-18X199-1	103	116-3X199-1	113	116-18X199-1	123	255-16-W	133	318-20	143
Y4	116-20X199-10	104	116-4X199-17	114	116-20X199-10	124	318-21X270-5	134	318-21X270-5	144
Y5	116-8X120-6	105	116-5X199-10	115	116-21-W	125	255-25 (8A)	135	318-22	145
Y6	120-9	106	116-8X120-6	116	116-23X199-8	126	318-9-W	136	318-24	146
Y7	120-12	107	116-10X199-2	117	116-24X199-16	127	318-10-W	137	270-2	147
Y8	120-12	108	116-11X199-10	118	318-22	128	318-11-W	138	270-3	148
Y9	120-17	109	116-12	119	[illegible]	129	318-13	139	270-7	149
Y10	120-12	110	116-13X120-8	120	270-3	130	318-15	140	270-8	150

	X6		X7		X8		X9		X10		
	Z2										
Y1	182-19 (7C)	151	120-23	161	116-14X120-8	171	255-10	181	318-16X198-7	191	Y1
Y2	182-21	152	116-1X199-9	162	116-15X120-8	172	255-16-W	182	318-17-W	192	Y2
Y3	116-1X199-9	153	116-3X199-1	163	116-18X199-1	173	255-16-W	183	318-20	193	Y3
Y4	116-4X199-17	154	116-4X199-17	164	116-20X199-10	174	255-21	184	318-21X270-5	194	Y4
Y5	116-5X199-10	155	116-5X199-10	165	116-21-W	175	255-25 (8A)	185	318-22	195	Y5
Y6	120-9	156	116-8X120-6	166	116-23X199-8	176	318-9-W	186	318-24	196	Y6
Y7	116-10X199-2	157	116-10X199-2	167	116-24X199-16	177	318-10-W	187	270-2	197	Y7
Y8	120-12	158	116-11X199-10	168	255-2-(10D)	178	318-11-W	188	270-3	198	Y8
Y9	120-17	159	116-12	169	[illegible]	179	318-13	189	270-7	199	Y9
Y10	120-19	160	116-13X120-8	170	[illegible]	180	318-15	190	270-8	200	Y10

	X1		X2		X3		X4		X5	
	Z3									
Y1	304-5-W	201	304-5-W	211	172-11 2b	221	288-12	231	312-7-W	241
Y2	129-6 2a	202	304-7-W	212	172-12x-9	222	288-15	232	312-8-W	242
Y3	129-7 2a	203	304-9	213	304-7-W	223	288-16	233	290-18 X 312-9	243
Y4	129-8	204	304-19	214	169	224	288-19	234	312-10-W	244
Y5	129-9 -W	205	304-19	215	288-1 2b	225	288-19	235	312-11-W	245
Y6	129-9	206	304-21	216	288-3	226	312-1-W	236	312-11-W	246
Y7	129-13	207	172-1 2a	217	288-5	227	312-2-W	237	312-12X288-10	247
Y8	129-20	208	172-1 2a	218	288-6	228	312-3x169	238	312-13	248
Y9	NC2477-14	209	NC2478-17	219	NC2192-2	229	312-4-W	239	312-14-W	249
Y10	304-1 2a	210	182 2a	220	288-9	230	312-5-W	240	312-15	250

	X6		X7		X8		X9		X10		
	Z3										
Y1	312-16	251	290-14	261	310-10	271	324-24	281	126-7	291	Y1
Y2	290-1-2C	252	290-18	262	310-13X290-15	272	324-16-W	282	126-11	292	Y2
Y3	290-2	253	290-21	263	310-14-W	273	324-17-2D	283	126-11	293	Y3
Y4	290-4	254	290-22	264	310-15-W	274	324-18	284	126-12	294	Y4
Y5	290-5	255	290-24	265	310-19	275	324-21-W	285	126-14	295	Y5
Y6	290-8	256	310-3-W	266	161-2c	276	324-22-W	286	126-15	296	Y6
Y7	290-9	257	310-3-W	267	310-24-W	277	324-24	287	126-17	297	Y7
Y8	290-11	258	270X310-4	268	324-17-2D	278	126-1	288	126-18-W	298	Y8
Y9	290-12	259	270-2c	269	324-12-W	279	290-5	289	126-18-W	299	Y9
Y10	290-13	260	310-8	270	324-12-W	280	126-4	290	126-20	300	Y10

	X1		X2		X3		X4		X5	
	Z4									
Y1	126-21-2D	301	114-9 X-114-20	311	307-14	321	307-9	331	240-14	341
Y2	114-3	302	114-21	312	307-20	322	307-17	332	240-15	342
Y3	114-6	303	114-23	313	307-21	323	307-18	333	240-18-240-10	343
Y4	114-8	304	307-2	314	307-22	324	307-21	334	240-20	344
Y5	114-6	305	307-3-W	315	307-3-W	325	240-5	335	240-23	345
Y6	114-4-114-11	306	307-5	316	307-24	326	240-6	336	240-24	346
Y7	114-12-126-18	307	307-7	317	306-4	327	240-10 X 240-18	337	216-1-3B	347
Y8	114-14	308	307-8	318	306-5	328	240-11	338	216-5	348
Y9	114-18	309	114-14	319	306-7	329	240-12	339	216-7X212-2 (12D)	349
Y10	114-19	310	307-13	320	306-10	330	240-13	340	216-6-3B	350

	X6		X7		X8		X9		X10		
	Z4										
Y1	216-7	351	185-1	361	279-25	371	185-22	381	110-14	391	Y1
Y2	185-1	352	216-23	362	185-3	372	185-23	382	110-17	392	Y2
Y3	216-10	353	184-2	363	185-4x185-2	373	185-24	383	173-25	393	Y3
Y4	185-1	354	185-1	364	185-7	374	110-3	384	110-22	394	Y4
Y5	216-13	355	184-11	365	185-12	375	110-6(	385	110-23	395	Y5
Y6	216-14	356	184-14	366	240-25	376	110-8	386	[illegible]	396	Y6
Y7	185-1	357	184-16X184-14	367	185-14	377	110-9	387	[illegible]	397	Y7
Y8	216-17	358	184-21X184-14	368	185-15	378	110-10	388	[illegible]	398	Y8
Y9	185-1	359	184-21X184-14	369	185-16	379	110-11	389	[illegible]	399	Y9
Y10	185-1	360	185-1	370	185-20-3C	380	110-12	390	[illegible]	400	Y10

	X1		X2		X3		X4		X5	
	Z5									
Y1	[illegible]	401	253-24-W	411	128-22	421	131-24	431	161-22	441
Y2	[illegible]	402	128-9	412	131-7	422	161-13-4A	432	161-24	442
Y3	[illegible]	403	128-10	413	131-9	423	161-14	433	268-7	443
Y4	253-1-W	404	253-24-W	414	131-12	424	161-15	434	268-9	444
Y5	253-1-W	405	253-24-W	415	271	425	161-16	435	268-10	445
Y6	253-6-W	406	128-15	416	131-12	426	161-17	436	268-11	446
Y7	112-11X253-12	407	128-16	417	131-18	427	161-18	437	268-12	447
Y8	253-19-W (3D)	408	128-18	418	131-19	428	161-19	438	268-13	448
Y9	253-20-W	409	128-20	419	131-12	429	161-19	439	268-14	449
Y10	253-21-W	410	128-21	420	131-22	430	161-21	440	268-16	450

	X6		X7		X8		X9		X10		
	Z5										
Y1	268-19	451	164-20	461	133-19	471	148-21	481	134-4	491	Y1
Y2	268-20	452	164-23	462	148-12	472	148-22	482	134-9	492	Y2
Y3	268-21	453	NC2365	463	148-13	473	134-10	483	134-10	493	Y3
Y4	268-19	454	NC2365	464	NC2365	474	NC2365	484	NC2365	494	Y4
Y5	164-1	455	NC2365	465	NC2365	475	NC2365	485	NC2365	495	Y5
Y6	164-2	456	NC2365	466	NC2365	476	NC2365	486	NC2365	496	Y6
Y7	164-16X164-10	457	NC2365	467	148-18	477	148-21	487	134-4	497	Y7
Y8	164-12	458	NC2365	468	148-13	478	164-20	488	164-20	498	Y8
Y9	164-18X164-10	459	NC2365	469	148-20	479	133-19	489	133-19	499	Y9
Y10	164-19 (4B)	460	NC2365	470	148-20	480	[illegible]	490	134-24	500	Y10

	X1		X2		X3		X4		X5	
	Z6									
Y1	132-7X132-5 (4C)	501	132-23	511	127-19	521	265-17	531	144-12	541
Y2	132-8	502	132-24	512	127-20-clf	522	265-18	532	144-13-W	542
Y3	132-10X132-4	503	127-1(4D)	513	127-21	523	265-19	533	144-14	543
Y4	132-15	504	127-7	514	127-24	524	265-20	534	144-15-W	544
Y5	132-16	505	127-8-clf	515	265-2	525	265-21	535	144-16	545
Y6	27-12X132-17 (4d	506	127-9-clf	516	127-1(4D)	526	265-22	536	44-18-144-19 (5A	546
Y7	132-18(4C)	507	127-11-clf	517	265-12	527	144-3-W	537	144-19-W	547
Y8	132-20-132-4	508	127-13-clf	518	265-13	528	144-6-W	538	144-20-W	548
Y9	132-20-132-9	509	127-14	519	265-15	529	144-8-W	539	144-22-W	549
Y10	1337-22	510	127-17	520	265-16	530	144-9-W	540	144-24	550

	X6		X7		X8		X9		X10		
	Z6										
Y1	271-1	551	271-7	561	147-17	571	273-12(5b)	581	137-13	591	Y1
Y2	271-7	552	271-15	562	244(5B)	572	273-13-W	582	137-14	592	Y2
Y3	271-7	553	271-17	563	269(5A)	573	273-17	583	137-15	593	Y3
Y4	271-4	554	271-24	564	147-24	574	273-18	584	137-16	594	Y4
Y5	271-7	555	147-1	565	273-1	575	273-18	585	273-19(4C,5B)	595	Y5
Y6	271-8	556	147-4-W	566	273-3	576	273-19	586	273-19	596	Y6
Y7	271-9	557	147-5	567	273-4	577	273-20	587	137-19	597	Y7
Y8	147-13	558	147-13	568	273-5	578	273-25	588	137-20	598	Y8
Y9	271-11	559	147-14	569	273-6	579	137-11	589	137-21	599	Y9
Y10	271-12	560	147-16	570	273-7	580	273-6	590	137-22	600	Y10

	X1		X2		X3		X4		X5	
	Z7									
Y1	137-23 (5B)	601	138-22-W	611	149-21	621	267-12	631	168-12	641
Y2	137-24	602	149-8 (5C)	612	149-22	622	267-17	632	168-14	642
Y3	138-3	603	149-14	613	149-23	623	267-18	633	168-15	643
Y4	138-7	604	149-15	614	149-24	624	287-18	634	168-16	644
Y5	246 (5B)	605	149-13	615	287-2	625	327-17	635	168-17	645
Y6	246 (5B)	606	149-14	616	287-4-W	626	168-7 (5D)	636	168-19	646
Y7	138-12	607	149-15	617	287-7-W	627	168-8	637	168-20	647
Y8	138-15-W	608	149-16	618	287-8	628	168-9X168-1	638	168-21	648
Y9	138-21	609	149-18	619	287-9	629	168-10	639	168-22	649
Y10	138-21	610	149-20	620	287-11-W	630	168-3X168-11	640	168-24	650

	X6		X7		X8		X9		X10		
	Z7										
Y1	243-12	651	173-5	661	106-10X106-11	671	106-22	681	155-11	691	Y1
Y2	243-12	652	173-8	662	106-11	672	106-24	682	155-13	692	Y2
Y3	173-13	653	173-12 (6A)	663	106-12	673	155-1	683	155-15	693	Y3
Y4	243-19	654-HB	173-13	664	106-14	674	155-3	684	155-16	694	Y4
Y5	243-20	655 ?	173-19	665	106-16	675	155-4	685	155-17	695	Y5
Y6	243-21	656	243-21	666	106-17	676	155-5	686	155-18	696	Y6
Y7	243-22	657	173-21	667	106-18	677	155-6	687	155-19	697	Y7
Y8	243-24	658	173-23	668	106-19	678	106-10	688	155-20	698	Y8
Y9	173-2	659	173-24	669	106-20	679	155-8	689	155-21	699	Y9
Y10	106-8	660	106-8	670	106-21	680	155-9	690	155-22	700	Y10

	X1		X2		X3		X4		X5	
	Z8									
Y1	155-23	701	272-17	711	285-6X5	721	285-6X5	731	154-6X285-21	741
Y2	155-24	702	272-17	712	163-11-W	722	285-7	732	285-21-W	742
Y3	272-5X272-15 (6B)	703	272-20	713	163-11-W	723	285-8	733	285-22	743
Y4	272-6 (6B)	704	272-6 (6B)	714	155-24	724	285-9	734	285-23	744
Y5	272-8X272-7	705	272-8X272-7	715	272-5X272-15 (6b)	725	285-21-W	735	285-24	745
Y6	272-9	706	272-9	716	154-6X285-21	726	285-18X2	736	154-2X285-2	746
Y7	272-12	707	272-12	717	154-5-W	727	163-6	737	154-4	747
Y8	272-13	708	272-13	718	272-16X15	728	285-15	738	154-4	748
Y9	272-14	709	163-4	719	285-4 (6C)	729	285-16	739	154-5-W	749
Y10	272-16X15	710	163-6	720	285-4	730	285-18X2	740	154-6	750

	X6		X7		X8		X9		X10		
	Z8										
Y1	154-7X103-13	751	154-21X103-14	761	103-13 (6D)	771	160-10	781	160-21	791	Y1
Y2	103-8X154-8	752	154-22-W	762	103-13 (6D)	772	160-11	782	160-22	792	Y2
Y3	154-10-W	753	154-24X103-6	763	160-11	773	160-11	783	160-23	793	Y3
Y4	154-11-W	754	154-12	764	160-21	774	160-22	784	118-2 (7A)	794	Y4
Y5	154-12	755	154-10-W	765	154-21X103-14	775	160-21	785	118-3	795	Y5
Y6	154-14X103-14	756	103-6	766	160-5	776	118-5	786	118-4	796	Y6
Y7	154-15	757	103-7-W	767	160-5	777	118-5	787	118-5	797	Y7
Y8	154-16X103-16	758	103-9	768	160-5	778	199-2X160-19	788	118-6	798	Y8
Y9	154-18X154-19	759	103-10	769	160-6	779	199-2X160-19	789	118-7	799	Y9
Y10	154-20	760	103-11	770	160-8	780	160-20	790	118-8	800	Y10

	X1		X2		X3		X4		X5	
	Z9									
Y1	118-20X9	801	118-21	811	284-1x19	821	280-11	831	152-13-W	841
Y2	118-10	802	118-22	812	284-20	822	280-12	832	152-14-W	842
Y3	118-11	803	118-10	813	284-23-W	823	280-14	833	276-25	843
Y4	118-13	804	280/4-4	814	284-24	824	280-16	834	152-16-W	844
Y5	118-14	805	280-4	815	280-5	825	280-18	835	152-17-W	845
Y6	118-16	806	284-5	816	280-6	826	280-19	836	182-2	846
Y7	118-17	807	284-7	817	280-7	827	280-20	837	182-3	847
Y8	118-18	808	249-25	818	280-8	828	280-22	838	182-5	848
Y9	118-19	809	284-13	819	280-8	829	280-24	839	182-18	849
Y10	118-20	810	284-18	820	280-10	830	[illegible]	840	182-18	850

	X6		X7		X8		X9		X10		
	Z9										
Y1	177-11	851	281-8	861	167-7	871	217-17-W	881	217-17-W	891	Y1
Y2	281-8	852	109-10-W	862	279-1(10A)	872	217-9	882	K169-14	892	Y2
Y3	177-14	853	109-10-W	863	279-1(10A)	873	217-12	883	K169-15 (wus?)	893	Y3
Y4	177-18	854	109-11	864	279-3	874	217-17-W	884	K169-16	894	Y4
Y5	177-23	855	109-12	865	K169-7(12A)	875	K169-7(12A)	885	194-13	895	Y5
Y6	281-2	856	109-13	866	279-5	876	K169-8	886	194-2	896	Y6
Y7	281-3	857	167-2	867	279-8	877	K169-9	887	K169-11	897	Y7
Y8	281-4	858	167-3	868	217-3 (10D)	878	K169-10	888	194-4	898	Y8
Y9	281-6	859	167-4	869	217-3 (10D)	879	K169-11	889	194-2	899	Y9
Y10	281-7	860	167-6	870	217-4	880	K169-12	890	K169-9	900	Y10

	X1		X2		X3		X4		X5	
	Z10									
Y1	270-9 (6B)	901	245-2	911	282-8-W	921	282-7	931	282-20	941
Y2	270-9 (8B)	902	245-4	912	245-19	922	282-8-W	932	282-22	942
Y3	270-11	903	245-6	913	245-20	923	282-9	933	282-23	943
Y4	270-12	904	245-8-W	914	245-21	924	282-11	934	282-24	944
Y5	270-14	905	245-8-W	915	245-22	925	282-13	935	278-1	945
Y6	270-16	906	245-10	916	245-24	926	282-14	936	278-2	946
Y7	270-19	907	245-13(8C)	917	282-1	927	282-14	937	282-6	947
Y8	270-20	908	245-16-W	918	282-3	928	282-17	938	278-5	948
Y9	270-22	909	245-16-W	919	282-4	929	282-18	939	278-6	949
Y10	270-24	910	245-17	920	282-6	930	282-19	940	278-7	950

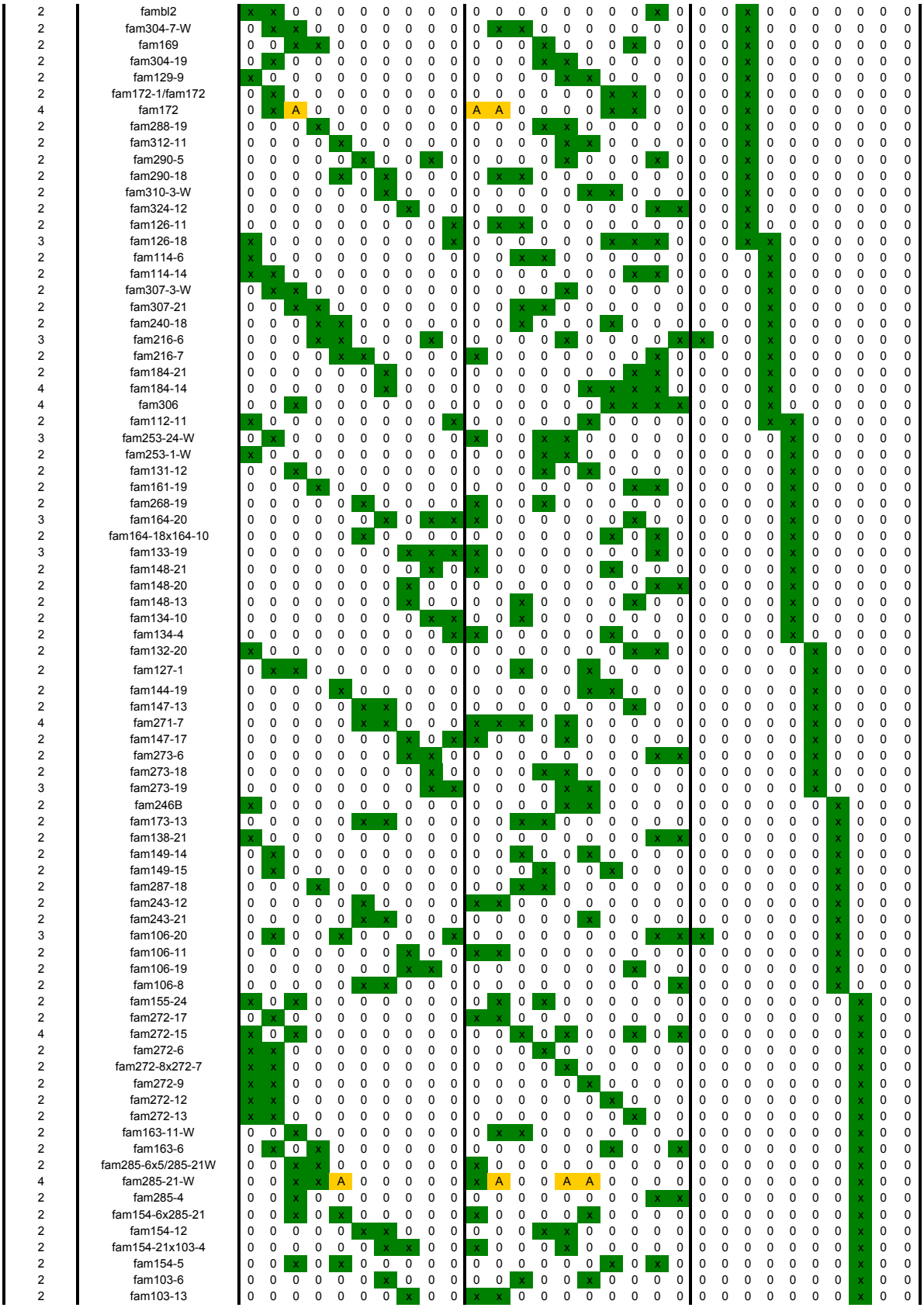
	X6		X7		X8		X9		X10		
	Z10										
Y1	278-24-W	951	278-21	961	283-12	971	286-7	981	286-23	991	Y1
Y2	278-10	952	278-22	962	283-13	972	286-8	982	286-21	992	Y2
Y3	278-13	953	278-24-W	963	283-17	973	286-9	983	286-22	993	Y3
Y4	278-14	954	283-2(8D)	964	278-22	974	286-11	984	250-1	994	Y4
Y5	286-16-W	955	283-3	965	278-24-W	975	286-12	985	250-2	995	Y5
Y6	278-16	956	283-4	966	283-22	976	286-13	986	250-3	996	Y6
Y7	278-17	957	278-14	967	283-24	977	286-14	987	250-4	997	Y7
Y8	278-18	958	283-8	968	286-1	978	286-15	988	250-5	998	Y8
Y9	278-19	959	283-9	969	286-3	979	286-16-W	989	250-6	999	Y9
Y10	278-20	960	283-11(9A)	970	286-5	980	286-19	990	177-5(9B)	1000	Y10

Supplementary Table 2: Oligo and adapter sequences

Oligo Name	Sequence (5'-3')
Bio-*MunI* adapter:	
Bio-*Mun* top	**BIOTIN**-CTCGTAGACTGCGTACG
Mun-bottom	AATTCGTACGCAGTC
MseI adapter:	
Mse-top	GACGATGAGTCCTGAG
Mse-bottom	TACTCAGGACTCAT
MunI-ACAC primer	AGACTGTGTACGAATTGACAC
MseI primer	GACGATGAGTCCTGAGTAA
Inverted repeat primers:	
NNNN-Iroutw primer	CATATATTAA**NNNN**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X1 primer	CATATATTAA**ACAC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X2 primer	CATATATTAA**ACAG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X3 primer	CATATATTAA**ACGA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X4 primer	CATATATTAA**ACGT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X5 primer	CATATATTAA**ACTC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X6 primer	CATATATTAA**ACTG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X7 primer	CATATATTAA**AGAC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X8 primer	CATATATTAA**AGAG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X9 primer	CATATATTAA**AGCA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool X10 primer	CATATATTAA**AGCT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y1 primer	CATATATTAA**AGTC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y2 primer	CATATATTAA**AGTG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y3 primer	CATATATTAA**ATCG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y4 primer	CATATATTAA**ATGC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y5 primer	CATATATTAA**CACA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y6 primer	CATATATTAA**CACT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y7 primer	CATATATTAA**CAGA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y8 primer	CATATATTAA**CAGT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y9 primer	CATATATTAA**CATC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Y10 primer	CATATATTAA**CATG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z1 primer	CATATATTAA**CGAT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z2 primer	CATATATTAA**CGTA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z3 primer	CATATATTAA**CTAC**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z4 primer	CATATATTAA**CTAG**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z5 primer	CATATATTAA**CTCA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z6 primer	CATATATTAA**CTCT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z7 primer	CATATATTAA**CTGA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z8 primer	CATATATTAA**CTGT**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z9 primer	CATATATTAA**GACA**GTAGCTCCGCCCCTG
NNNN-IRoutw-pool Z10 primer	CATATATTAA**GACT**GTAGCTCCGCCCCTG
extension primer	CATATACAATTGGACGATGAGTCCTGAGTAA
GS20 *MunI* adapter B:	
GS20-*Mun*-top	CCTATCCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTCTCAG
GS20-*Mun*-bottom	AATTCTGAGACACGCAACAGGGGATAGGCAAGGCACACAGGGGA
GS20 *MseI* adapter A:	
GS20-*Mse*-top	CCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTCTCAG
GS20-*Mse*-bottom	TACTGAGACAGGGAGGGAACAGATGGGACACGCAGGGATGAG
Amplification adapter primer A & B:	
BioTEG-adapter primer B	**BIOTIN-TEG**-CCTATCCCCTGTGTGCCTTG
adapter primer A	CCATCTCATCCCTGCGTGTC
sequence primer A	CCATCTGTTCCCTCCCTGTC
NAM screening:	
IRoutw	GAATTCGCTCCGCCCCTG
PhNAM_REV	TTTGTTGAGGTCAACTTCAGCAATG
PhNH26_REV	CTTTATTGGCCTAGTGAGGAGC
PhNH3_REV	ATACTCGTGCATTATCCAGTTAGTC
PhNH10_REV	ATCTGATAACTGCTCTGAAACTGTG
PhNH12-LIKE_REV	CCCATTTGGATACTTTCTATCTC
PhNH15_FW	TCCAACAGGTCTACGGACTAACC
PhNH16/PhCUC3_FW	CAAGCAGTTGGACAAAATTGTGG
PhNH17_FW	ACTACAAGCAAGATCAGCGGTTAC
PhNH22_REV	ACACAAGTGTTTTCTTCATGCCAAC
PhNH24_REV	TGGCCATGGGGAGCTCTTCC
PhNH27_REV/PhNH25_REV	CCTGTTGCCTTCCAGAATCCT

Supplementary Table 3: Small family search profiles

Total indiv.	FamilyName	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
2	fam229-7	x	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam229-8/fam229-10	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	fam229-10	x	A	0	0	0	x	0	0	0	0	0	x	0	A	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam229-9-w	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	fam229-11-W	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	fam229-16	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	x	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam229-17	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	0
4	famK218-6	x	x	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	x	x	0	0	0	0	0	0	0	0	0
2	fam229-23-W	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam229-24-W	0	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam113-7-W/fam113-4-W	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	fam113-4-W	0	A	x	0	0	0	0	x	0	0	0	0	0	0	0	x	A	0	0	0	x	0	0	0	0	0	0	0	0	0
3	fam113-3-W	0	0	x	0	0	0	x	x	0	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	0
5	fam113-2-W	x	0	x	0	0	0	x	x	x	0	0	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0	0
3	famK218-8	0	x	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	0
2	famK218-7	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	x	x	0	x	0	0	0	0	0	0	0	0	0
3	famK262-9-W	0	x	0	0	x	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK262-3-W	0	0	0	0	x	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK262-2	0	0	0	0	x	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK262-1	0	0	0	0	x	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
4	fam212-10-W	0	x	0	0	0	0	x	x	0	0	0	0	0	0	x	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam212-18	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam216-6-12D	0	0	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	famK234	0	0	x	0	0	0	0	0	0	0	0	0	x	x	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK111-6-W	0	0	0	x	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK111-5-W	0	0	0	x	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	famK111-3-W	0	0	0	x	0	0	0	x	x	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	famK111-1-W	0	0	0	x	0	0	0	x	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
4	famK111-7-W	0	0	x	x	0	x	x	x	0	0	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0
4	fam198-15/212-5	0	0	0	x	x	0	0	0	x	0	0	0	0	0	0	x	x	0	x	0	x	0	0	0	0	0	0	0	0	0
2	famE210-9	0	0	0	0	x	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famE210-5	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0	0
2	famK270-6-W	0	0	x	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
3	famK272?-7-W	0	0	x	0	0	0	x	x	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
2	fam255-25	0	0	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam318-21x270-5	0	0	0	x	x	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam270-2	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam270-7	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0
2	fam270-8/270-3	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0
3	fam270-3	0	0	A	0	x	0	0	0	0	x	0	0	0	0	0	0	0	A	0	x	0	x	0	0	0	0	0	0	0	0
3	fam318-22	0	0	x	0	x	0	0	0	0	x	0	0	0	0	x	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-14x120-8	x	0	x	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-15x120-8	0	0	x	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-13x120-8	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0
3	fam116-23x199-8	x	0	x	0	0	0		x	0	0	0	x	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-18x199-1	x	0	x	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
5	fam199-1	x	x	x	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-20x199-10	x	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-8x120-6	x	x	0	0	0	0	x	0	0	0	0	0	0	0	x	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam120-9	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0
4	fam120-12	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	x	0	x	0	x	0	0	0	0	0	0	0	0
2	fam120-23	0	x	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-1x199-9	0	x	0	0	0	x	x	0	0	0	0	x	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-3x199-1	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-4x199-17	0	x	0	0	0	x	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-5x199-10	0	x	0	0	0	x	x	0	0	0	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam116-10x199-2	0	x	0	0	0	x	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-11-x199-10	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-12	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0
2	fam116-13x120-8	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0
2	fam116-21-W	0	0	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam116-24x199-16	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam295-8	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0
3	fam318-16x198-7	0	0	0	x	x	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam318-20	0	0	0	x	x	0	0	0	0	x	0	x	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
3	fam295-18-W	0	0	0	x	0	0	0	0	x	0	0	x	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam318-9-W	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam318-10-W	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam318-11-W	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0
2	fam318-13	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	x	0	0	0	0	0	0	0	0
2	fam318-15	0	0	0	x	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0
2	fam318-17-W	0	0	0	0	x	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam318-24	0	0	0	0	x	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	x	0	0	0	0	0	0	0	0
2	fam304-5-W	x	x	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0



3	fam103-14	0	0	0	0	0	x	x	x	0	0	x	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
3	fam160-11	0	0	0	0	0	0	0	x	x	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
3	fam160-21	0	0	0	0	0	0	0	x	x	x	x	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam160-22	0	0	0	0	0	0	0	0	x	x	0	x	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	199-2x160-19	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	x	0	0
3	fam160-5	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	x	x	x	0	0	0	0	0	0	0	0	0	0	x	0	0
3	fam118-5	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam118-10	x	x	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0
2	fam118-20	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	x	0	0
4	fam152-W	0	0	0	0	x	0	0	0	0	0	x	x	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam182-18	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	x	0	0
2	fam281-8	0	0	0	0	0	x	x	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam109-10-W	0	0	0	0	0	0	x	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam279-1	0	0	0	0	0	0	0	x	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
3	fam217-17-W	0	0	0	0	0	0	0	0	x	x	x	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0
2	fam217-3	0	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	x	0	0
4	fam194	0	0	0	0	0	0	0	0	0	x	0	0	0	0	x	x	0	x	x	0	0	0	0	0	0	0	0	0	x	0	0
2	fam270-9	x	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x
2	fam245-8-w	0	x	0	0	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x
2	fam282-8-w	0	0	x	x	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x
2	fam282-14	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0	0	0	0	0	x
3	fam278-24	0	0	0	0	0	x	x	x	0	0	x	0	x	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x
2	fam278-22	0	0	0	0	0	0	x	x	0	0	0	x	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x
2	fam286-16-W	0	0	0	0	0	x	0	0	x	0	0	0	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	x

Supplementary Table 4: Large family search profiles

total indiv.	Family name	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
23	F229	>1	x	0	0	0	>1	x	0	0	0	x	x	x	x	x	x	x	x	x	0	x	0	0	0	0	0	0	0	0	0
13	F113-W	x	x	A	0	0	0	x	A	x	0	0	0	0	0	0	A	A	A	x	0	x	0	0	0	0	0	0	0	0	0
9	FK218	x	x	0	x	x	0	0	0	x	0	0	0	0	0	0	0	0	A	x	A	x	0	0	0	0	0	0	0	0	0
9	FK262	0	x	0	0	A	0	0	0	0	A	x	x	A	A	0	0	0	0	0	0	x	0	0	0	0	0	0	0	0	0
9	F212	0	x	0	x	x	0	x	x	x	0	0	0	0	0	A	A	A	0	x	0	>1	0	0	x	0	0	0	0	0	0
14	FK111	0	0	x	A	0	x	x	x	A	0	x	x	A	A	0	0	0	0	0	x	x	0	0	0	0	0	0	0	0	0
6	FE210	0	0	0	0	x	0	0	0	0	>1	0	0	0	0	x	x	x	x	0	0	x	0	0	0	0	0	0	0	0	0
7	F198	0	0	0	A	A	0	0	0	x	x	x	0	0	0	0	x	x	0	x	0	>1	x	0	0	0	0	0	0	0	0
53	F199/116/120	x	x	x	0	0	x	x	x	x	0	x	x	x	x	x	A	A	x	x	x	0	>1	0	0	0	0	0	x	0	0
7	F120-8	x	x	x	0	0	0	x	x	0	0	x	x	0	0	0	0	0	0	0	x	0	x	0	0	0	0	0	0	0	0
8	F199-10	x	x	x	0	0	x	x	x	0	0	0	0	0	x	x	0	0	x	0	0	0	x	0	0	0	0	0	0	0	0
26	F318	0	0	x	x	A	0	0	0	x	A	x	x	x	x	x	x	x	x	x	x	0	x	0	0	0	0	0	0	0	0
9	F295	0	0	x	x	0	0	0	A	A	0	x	x	x	x	0	0	0	x	x	x	0	x	0	0	0	0	0	0	0	0
9	F304	x	>1	x	0	0	0	0	0	0	0	x	x	x	x	x	x	0	0	0	x	0	0	x	0	0	0	0	0	0	0
7	F129	x	0	0	0	0	0	0	0	0	0	0	x	x	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0	0	0
11	F288	0	0	A	A	x	0	0	0	0	0	x	x	x	x	x	x	x	x	0	x	0	0	x	0	0	0	0	0	0	0
16	F312	0	0	0	x	>1	x	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	x	0	0	0	0	0	0	0
17	F290	0	0	0	0	x	>1	x	x	x	0	x	x	x	x	x	x	x	x	x	x	0	0	x	0	0	0	0	0	0	0
10	F310	0	0	0	0	0	0	x	x	0	0	x	x	x	x	x	x	x	x	0	x	0	0	x	0	0	0	0	0	0	0
10	F324	0	0	0	0	0	0	0	x	>1	0	x	x	x	x	x	x	x	x	x	x	0	0	x	0	0	0	0	0	0	0
8	F185-1	0	0	0	0	0	x	x	0	0	0	x	x	0	x	0	0	x	0	x	x	0	0	0	x	0	0	0	0	0	0
6	F184	0	0	0	0	0	0	x	0	0	0	0	0	x	0	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0	0
14	F126	x	0	0	0	0	0	0	0	x	>1	x	x	x	x	x	x	x	x	x	x	0	0	>1	x	0	0	0	0	0	0
13	F114	>1	x	0	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	x	0	0	0	0	0	0
16	F307	0	A	x	A	0	0	0	0	0	0	x	x	x	x	x	x	x	x	0	x	0	0	0	x	0	0	0	0	0	0
4	F306	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0	x	x	x	x	0	0	0	x	0	0	0	0	0	0
13	F240	0	0	0	A	A	0	0	x	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	x	0	0	0	0	0	0
12	F216	0	0	0	x	A	A	x	0	x	0	x	x	x	0	x	x	x	x	x	x	x	0	0	>1	0	0	0	0	0	0
19	F185	0	0	0	0	0	x	x	>1	x	0	x	x	x	x	x	0	x	x	x	x	0	0	0	x	0	0	0	0	0	0
6	F279	0	0	0	0	0	0	0	x	0	0	x	x	x	x	0	x	0	0	0	0	0	0	0	x	0	0	0	0	>1	0
11	F110	0	0	0	0	0	0	0	0	x	x	x	x	0	x	x	x	x	x	x	x	0	0	0	x	0	0	0	0	0	0
9	F112	x	0	0	0	0	0	0	0	0	A	x	x	x	0	0	x	x	x	x	x	0	0	0	A	x	0	0	0	0	0
10	F253	>1	x	0	0	0	0	0	0	0	0	x	0	0	x	x	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0
8	F128	0	>1	x	0	0	0	0	0	0	0	x	x	x	0	0	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0
9	F131	0	0	>1	x	0	0	0	0	0	0	x	x	x	x	0	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0
14	F271	0	0	x	0	x	>1	x	0	0	0	x	x	x	x	x	x	x	0	x	x	0	0	0	0	x	>1	0	0	0	0
12	F161	0	0	0	>1	x	0	0	x	0	0	x	x	x	x	x	x	x	x	x	x	0	0	x	0	>1	0	0	0	0	0
12	F268	0	0	0	0	>1	x	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0
10	F164	0	0	0	0	0	>1	x	0	x	x	x	x	0	0	x	x	x	x	x	x	0	0	0	0	x	0	0	0	0	0
17	FNC2365	0	0	0	0	0	0	>1	x	x	x	0	0	x	A	A	A	x	x	x	x	0	0	0	0	x	0	0	0	0	0
6	F134	0	0	0	0	0	0	0	0	x	>1	x	x	x	0	0	0	x	0	0	x	0	0	0	0	x	0	0	0	0	0
9	F148	0	0	0	0	0	0	0	x	x	0	x	x	x	0	0	0	x	x	x	x	0	0	0	0	x	0	0	0	0	0
11	F132	>1	x	0	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	x	0	0	0	0
14	F127-clf	x	>1	x	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	x	0	0	0	0
11	F265	0	0	x	x	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	x	0	0	0	0
14	F144	0	0	0	x	>1	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	x	0	0	0	0
10	F147	0	0	0	0	0	0	>1	x	0	x	x	0	0	x	x	x	x	x	x	x	0	0	0	0	0	x	0	0	0	0
17	F273	0	0	0	0	0	0	0	x	x	x	x	x	x	x	A	A	x	x	x	x	0	0	0	0	0	x	0	0	0	0
11	F137	x	0	0	0	0	0	0	0	x	>1	x	x	x	x	0	0	x	x	x	x	0	0	0	0	0	>1	x	0	0	0
7	F138	>1	x	0	0	0	0	0	0	0	0	x	0	x	x	0	0	x	x	x	x	0	0	0	0	0	0	x	0	0	0
15	F149	0	>1	x	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	x	0	0	0
10	F287	0	0	x	x	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	x	0	0	0
11	F173	0	0	0	0	0	x	>1	0	0	x	x	x	x	x	x	0	x	x	x	0	0	0	0	x	0	0	>1	0	0	0
15	F168	0	0	0	x	>1	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	x	0	0	0
8	F243	0	0	0	0	0	>1	x	0	0	0	x	x	0	x	x	x	x	x	0	0	0	0	0	0	0	0	x	0	0	0
19	F106	0	x	0	0	x	x	x	>1	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	>1	0	0	0
20	F155	x	0	x	0	0	0	0	0	x	>1	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	>1	x	0	0
21	F272	A	A	x	0	0	0	x	x	0	0	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	>1	0	0
5	F163	0	x	x	x	0	0	0	0	0	0	0	x	x	0	0	0	x	0	x	x	0	0	0	0	0	0	0	x	0	0
19	F285	0	0	x	>1	x	0	0	0	0	0	A	A	A	A	x	x	0	x	x	x	0	0	0	0	0	0	0	x	0	0
24	F154	0	0	x	0	x	>1	x	x	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	x	0	0
14	F103	0	0	0	0	0	x	>1	x	0	0	x	x	x	0	x	x	x	x	x	x	0	0	0	0	0	0	0	x	0	0
19	F160	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	x	0	0
22	F118	>1	x	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	x	>1	0
9	F284	0	x	x	0	0	0	0	0	0	0	x	x	x	x	0	x	x	x	x	x	0	0	0	0	0	0	0	0	x	0
16	F280	0	x	x	x	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	x	0
7	F281	0	0	0	0	0	>1	x	0	0	0	x	x	0	0	0	x	x	x	x	x	0	0	0	0	0	0	0	0	x	0
8	F182	0	x	0	0	>1	x	0	0	0	0	x	x	0	0	0	x	x	x	x	x	0	x	x	0	0	0	0	0	>1	0
5	F177	0	0	0	0	0	>1	0	0	0	x	x	0	A	A	A	0	0	0	0	x	0	0	0	0	0	0	0	0	>1	x
5	F109	0	0	0	0	0	0	x	0	0	0	0	x	x	x	x	x	0	0	0	0	0	0	0	0	0	0	0	0	x	0
5	F167	0	0	0	0	0	0	>1	x	0	0	x	0	0	0	0	0	A	A	A	A	0	0	0	0	0	0	0	0	x	0
14	FK169	0	0	x	x	0	0	0	x	A	A	0	x	x	x	x	x	x	x	x	x	0	0	x	0	0	0	0	0	>1	0
8	F217	0	0	0	0	0	0	0	x	x	x	x	x	x	x	0	0	0	x	x	x	0	0	0	0	0	0	0	0	x	0
15	F245	0	>1	x	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	x
20	F282	0	0	x	>1	x	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	x
20	F278	0	0	0	0	x	>1	x	x	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	x
11	F283	0	0	0	0	0	0	x	x	0	0	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	x
17	F286	0	0	0	0	0	x	0	x	>1	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	x
6	F250	0	0	0	0	0	0	0	0	0	x	0	0	0	x	x	x	x	x	x	0	0	0	0	0	0	0	0	0	0	x
26	F270	>1	0	x	x	x	0	x	x	0	x	x	x	x	x	x	x	x	x	x	x	x	>1	x	0	0	0	0	0	0	>1
29	F304/306/307	x	x	x	x	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	0	0	>1	>1	0	0	0	0	0	0
105	F-NB1513(190-239)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	>1	x	0	x	0	0	0	x	x	0

Supplementary Table 5: Database lines representing assigned NAC insertions

						X-dimension										Y-dimension										Z-dimension												
alleles	Annotation	id	X	Y	Z	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	-	Cl. Size	seq…
phnh3-1	#912	15756	1	1	1	0	3	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2		8	CCCAGGACCTTAGGGATCAAGAAGGCACTAGT
phnh3-2	F154/103	19127	2	1	1	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	5	0	0		14	CCAAGAGATAGAAAATACCCAAATGGTTCACGG
phnh10-1	#572	16952	1	1	1	0	0	0	0	0	0	0	5	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0		10	ACCGAAAGCCAGAGATTCTACCACCGCTACCA(
phnh17-1	F284	36289	2	3	1	0	2	8	0	0	0	0	0	0	0	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	20	ATAACTACTGGAACTGGCAAGGTGGTTGTTGTC
phnh22-1	#205	13798	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0		6	GAAGGCAACAGGACTTGACAAGCAAATAGTGA(
phnh23-1	#247	30744	1	1	1	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0		10	GCAGCTAAGGCGACGTTTGGAGAACAAGGAAT
phnh24-1	F167	7367	2	3	1	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	4	3	2	0	0	0	0	0	0	0	0	5	0		20	GGGAGGGGACAAATGCATAAGGGAACAGCTTC
phnh25-1	#322	45951	1	1	1	0	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0		6	AGCAGCAGGATTTTGGAAGGCAACAGGAAGAG
phnh26-1	F280/84	25950	2	2	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0		6	TACTCAACAAGTGAATCATTCTGGCATGGTGAG