Supplementary Methods:

Meiotic gene evolution: can you teach a new dog new tricks?

Authors:
Lloyd AH[1,2], Ranoux M[3], Vautrin S[4], Glover N[3,5], Fourment J[4], Charif D[1,2], Choulet F[3,5], Lassalle G[6], Marande W[4], Tran J[1,2], Granier F[1,2], Pingault L[3,5], Remay A[3,5], Marquis C[1,2], Belcram H[7], Chalhoub B[7], Feuillet C[3,5], Bergès H[4], Sourdille P[3,5], Jenczewski E[1,2].

Author affiliations:

[1] INRA, UMR1318, Institut Jean-Pierre Bourgin, RD10, F-78000 Versailles, France
[2] AgroParisTech, Institut Jean-Pierre Bourgin, RD10, F-78000 Versailles, France
[3] INRA UMR1095 Genetics Diversity and Ecophysiology of Cereals, F-63039 Clermont-Ferrand, France.
[4] INRA UPR 1258, Centre National des Ressources Génomiques Végétales, Castanet-Tolosan, France
[5] University Blaise Pascal, UMR1095 Genetics Diversity and Ecophysiology of Cereals, F-63039 Clermont-Ferrand, France.
[6] INRA, UMR1349-IGEPP, Le Rheu Cedex, France.
[7] INRA-CNRS-UEVE, Unité de Recherche en Génomique Végétale, Evry, France.

Identification of genes sharing common ancestry with known meiotic genes

We first reviewed and established a list of 65 genes that have been experimentally shown to be involved in plant meiosis (Table S2). Based on the phenotype of mutants, they were shown to encompass a wide range of processes such as meiotic recombination, the control of cell cycle, sister chromatid cohesion and release, condensation or chromosome axes. 31 (48%) of the 65 genes have no known extra-meiotic function; the remainder have demonstrated roles in somatic cells in addition to their meiotic function.

We then used PLAZA 2.5 (Van Bel et al. 2012), Phytozome 7.0 (Goodstein et al. 2012) and EnsemblPlants (http://plants.ensembl.org/index.html) to select and retrieve comprehensive sets of homologous sequences (i.e. that share common ancestry with each of the 65 known meiotic genes) among the 18 angiosperm genomes of our survey. As these web resources use different (combinations of) methods to find and group homologous genes into families (see Table S23), we combined and compared their output in order to ascertain that no homologous gene was omitted. In addition to a common set of 15 species, PLAZA 2.5, Phytozome 7.0 and EnsemblPlants also contain complementary species representing WGD events that are not shared with any of the species present in the common set. The genomes of *Brassica rapa* (Chinese cabbage; X.X. Wang et al. 2011), *Solanum Lycopersicon* (Tomato; The Tomato Genome Consortium 2012), *Solanum tuberosum* (Potato; Xu et al. 2011) and *Musa acuminata* (Banana; D'Hont et al. 2012) are included in the Ensembl database while Phytozome provides access to the genomes of *B. rapa* and *Gossypium raimondii* (Diploid cotton; K. Wang et al. 2012). To be absolutely sure that the retrieved lists of homologues are exhaustive, we carried out BLASTP and BLASTN searches against the predicted proteome / scaffolds of these five species using the web-accessible BLAST tools provided for every (group of) species (link provided below). BLASTP and BLASTN searches were performed using *Arabidopsis thaliana* CDS and protein sequences as queries and the default parameters of the web servers. This procedure was straightforward for *B. rapa*, which is the closest relative to Arabidopsis (divergence time <20 million years ago), enabling us to use a lower e-value cut-off (<1e-95) than for the other species (roughly <1e-70 for cotton; <1e-50 for tomato and potato; <1e-30 for banana). For tomato and potato, we also confirmed our BLASTP and BLASTN results by searching the gene families defined by the Tomato Genome Consortium (2012). For banana, we reiterated the BLASTP and BLASTN searches using rice, or maize, CDS and protein sequences as queries. Finally the list of best BLAST hits were examined on a gene-by-gene basis to discard the sequences that share a very conserved domain with, but are not homologous to a given meiotic gene/protein (e.g. HORMA domain of ASY1).

Most frequently the gene families in PLAZA 2.5, Phytozome 7.0 and EnsemblPlant were completely congruent. Whenever a difference was noted, the discordant gene(s) was (were) carefully inspected. In most cases, discordant genes proved to be partial copies, which were retrieved by one web service but not the others. In a few other cases, which concerned gene families with old duplicates and numerous members (such as *Mei2-like*, *RPA* or *CDK* genes), inconsistencies arose when a gene was assigned to different clusters by the different tools (e.g. Glyma09g07850 clusters together with *AtRPA1* in EnsemblPlants but not in Plaza 2.5). Finally, in one occasion (Shugoshin family), sequence diversity between species was so large that PLAZA 2.5 clustering was unable to reconstruct the entire gene family; in this instance we used data from the literature (SGO was first described in maize by Hamant et al. 2005, then in rice by M. Wang et al. 2011) and other web services to retrieve all the genes related to SGO in plants. In all cases, all duplicates identified within a genome (whether they are supported by all or only one method) were considered for further analyses (see Assignment of duplicates to WGDs below).

Once a list of genes sharing common ancestry with a known meiotic gene was established, we examined every gene model in every species in order to identify and manually curate those erroneously annotated: i.e. single coding sequence (CDS) annotated as a series of linearly arrayed short CDS corresponding to the successive parts of a given protein, or distinct but adjacent genes models that were first annotated as a single CDS (see comments in Tables S5-S15).

Links:

BRAD database: http://brassicadb.org/brad/index.php

Sol Genomics Network: http://solgenomics.net/tools/blast/index.pl

Banana genome Hub: http://banana-genome.cirad.fr/index.html

Assignment of duplicates to Whole Genome Duplications

Gene pairs within a genome were considered to have arisen from a given WGD if a) they were located in syntenic blocks arising from that WGD, b) had a number of synonymous substitutions per synonymous site (Ks) consistent with the genome-wide average for that WGD and c) if their position in the phylogenetic tree is consistent with such an origin.

Synteny

Where possible, we used the WGDotplot applet of PLAZA 2.5 (Van Bel et al. 2012) and/or the SynMap applet of the CoGe platform (Lyons and Freeling 2008) to examine whether duplicated genes within a genome are located within collinear sets of genes. A hyperlink is provided in Tables S5-S15 for every pair of duplicates that reside in syntenic blocks. Syntenic and non-syntenic orthologs between *A. thaliana* and each of the three *B. rapa* subgenomes (Cheng et al. 2012; Tang et al. 2012) were identified using the BRAD data mining tools (Cheng et al. 2011). For cereals, we used MCScanX (Y. Wang et al. 2012) to investigate whether duplicates are located within collinear sets of genes, and determined whether these regions originate from the same ancestral chromosomes using the information published in Murat et al. (2010). In the Musa lineage, three whole genome duplication events, denoted as alpha, beta and gamma (from the most recent event to the oldest) were reported by D'Hont et al (2012). Duplicate copies of meiotic genes deriving from alpha/beta WGDs were identified by screening the beta ancestral blocks. For each meiotic gene we checked if a copy could be found in a duplicated region involved in a beta ancestral block. Musa analysis was undertaken by O. Garsmeur, CIRAD.

Ks

Ks values were estimated for each duplicate gene pair using the Yang and Nielsen method implemented in the yn00 program in the PAML package (Yang 2007). Protein sequences were first aligned using CLUSTALW (Larkin et al. 2007) and the protein alignments were used to guide coding sequence alignments by PAL2NAL (Suyama et al. 2006). Ks estimates are provided in Tables S5-S15 for every pair of duplicates.

Phylogenetic analysis

We used phylogenetic trees of gene families provided by PLAZA 2.5 except when these were not expected to be relevant e.g. when based on very short alignments, or when genes/species were missing from the PLAZA 2.5 database. In these instances we individually constructed phylogenetic trees using Phylogeny.fr (Dereeper et al. 2008; Dereeper et al. 2010). In both cases, phylogenetic trees were constructed with PhyML (Guindon and Gascuel 2003) based on multiple amino acid sequence alignments generated by MUSCLE (Edgar 2004). As WGDs can be unique to some lineages (e.g. the most recent WGDs in *B. rapa* and *G. max*, respectively) or conversely be shared by sister lineages (the α WGD affected both the *A. thaliana* and *B. rapa* genomes, the B WGD affected both the *G. max* and *Medicago truncatula* genomes), we mapped all duplication events relative to diagnostic speciation events in order to determine whether the duplication occurred before or after the species diverged.

Combining these three criteria was not possible for all duplicates because, for example, some copies have still to be anchored to chromosomes. Despite this only 13 (1.3%) of the ~1000 meiotic gene duplication events observed could not be attributed indisputably to polyploidy or non-polyploidy events. In addition, as meiotic genes show lower duplicate retention than genome average, we considered all ambiguous meiotic gene pairs as having arisen through WGD. This approach minimised any artificial bias toward higher meiotic gene loss.

Determining whether a duplicate pair arose through WGD becomes more difficult the older the WGD. This is due to Ks saturation, gene movement and progressive disruption of syntenic blocks. Whenever initial phylogenetic analyses suggested duplication predating the monocot-dicot divergence, we carried out more rigorous phylogenetic analyses for confirmation. EST/CDS sequences from basal angiosperms (*Aristolochia*, *Liriodendron*, *Nuphar*, *Amborella*) and additional monocots (*Phoenix dactylifera*) were identified by BLASTN searches using the AAGP and CoGe (Lyons and Freeling 2008) databases. These sequences were aligned at the amino acid level with the other members of a given OGs using either ClustalW or MUSCLE. Alignments of the corresponding DNA sequences were then guided by the amino acid alignments using PAL2NAL and adjusted by eye in BioEdit as necessary. These large alignments were then trimmed by removing poorly aligned regions. Individual sequences were subsequently removed from the alignment if the sequence contained less than 50% of the total alignment length. Maximum likelihood analyses were then conducted using PhyML (Guindon and Gascuel 2003), searching for the best maximum likelihood tree with the GTR model (determined using jModelTest 2.1; Darriba et al. 2012). Duplicates originating prior to the monocot dicot divergence were not further considered in our analyses.

Links:

Ancestral Angiosperm Genome Project (AAGP): http://ancangio.uga.edu/content/aagp-home

Comparative Genomics: http://genomevolution.org/CoGe/

Duplicate gene loss modelling

We used a likelihood-ratio test to determine which of two models best fit the observed duplicate gene loss data. The survival curve $S(Ks)$ describing duplicate loss for the two

population model, in which a given percentage of the duplicates (p) have a short half-life ($Ks_{1/2:S}$) while the remainder ($1\text{-}p$) have a long half-life ($Ks_{1/2:L}$), can be produced from the following equation.

$$S(Ks) = p \times 2^{Ks/Ks_{1/2:S}} + (1-p) \times 2^{Ks/Ks_{1/2:L}}$$

Constraining this model (p = 0 or 1) gives the single population exponential decay model in which decay occurs at a constant rate.

$$S(Ks) = 2^{Ks/Ks_{1/2}}$$

After maximum-likelihood estimation under the two models, twice the difference in log likelihood between the single and two-population models was compared to a chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters, i.e. 2. Gene-loss data for maize, (a clear outlier, Figure 1A) were removed from the analyses.

The decay rates (hazard functions) of the survival curves were calculated using the following formula.

$$\lambda(Ks) = -\frac{S'(Ks)}{S(Ks)}$$

Test of molecular evolution

We searched for evidence of acceleration in the rate of non-synonymous substitution using phylogeny-based approaches. Global nonsynonymous (Ka) to synonymous (Ks) nucleotide substitution ratios were estimated for all pairs of gene duplicate using the method of Yang and Nielsen (2000), as implemented in PAML 4 (Yang 2007). Ka/Ks estimates are provided in Tables S5-S15 for every pair of duplicates.

Tajima's Relative Rate test (Tajima 1993), as implemented in MEGA version 5.05 (Tamura et al. 2011) was used to determine if one of the retained copies was evolving faster than the other one. The designated outgroup sequence is a single-copy ortholog sampled in the genome of a closely-related species. We used genes from the following species as outgroups in the analysis: ***Arabidopsis thaliana:*** *Carica Papaya*; ***Glycine max:*** *Medicago truncatula* or *Lotus japonicus* or *Fragaria vesca* (as a final resort); ***Populus trichorcarpa:*** *Ricinus communis* or *Manihot esculenta*; ***Malus domestica:*** *Fragaria vesca*; ***Zea mays:*** *Sorghum bicolor;* ***Brassica rapa:*** *Arabidopsis thaliana*. For gene duplicates showing a putative asymmetric rate of evolution, sequence alignments were manually curated to ensure that the different evolutionary rates observed were not an artificial consequence of differences in codon prediction in the two gene annotations.

For genes originating from ancient WGDs, we searched for changes in selective pressure on particular branches or on a fraction of amino acids in a branch using the branch or branch-site models of PAML 4. Likelihood ratio (LR) test were carried out to test whether (some of) these models provided a better fit than the null model in which all branches (or all sites) have the same ω.

Analysis of gene fractionation

We looked for signs of gene fractionation using the CoGe comparative genomics platform (Lyons and Freeling 2008). For each gene pair, we used CoGeBLASTN searches to retrieve genomic sequences from *A. thaliana*, *Brassica rapa*, *Glycine max* and *Medicago truncatula* (when assessing fractionation in *G.max* only) from the CoGe database. We then visually scanned all DNA-DNA alignments implemented in the GEvo to look for gaps between the *A. thaliana* / *M. truncatula* reference sequence and any of the two or three retained genes in *B. rapa* and *G. max.* We used both BLASTn and BLASTz (Schwartz et al. 2003) comparisons for pairwise combinations.

Analysis of gene retention by function

Gene ontology enrichment analysis was performed using the AmiGO web application (http://amigo.geneontology.org; Carbon et al. 2009). As broad GO groupings can limit the identification of biased gene retention in particular physiological or biochemical contexts (Coate et al. 2011), we split genes into 14 meiotic functional subclasses (Table S17) in order to further investigate duplicate retention. Statistical analysis used chi-square tests with Bonferroni correction for multiple testing.

BAC screening

To ensure the correct identification of all members of a gene family in wheat (for which only low coverage sequence is available) and for oilseed rape (for which no sequence is available) we performed BAC library and PCR screening, focussing on 19 meiotic recombination genes that we identified as frequently reverting to a single copy following WGD.

a) Organisation of *Brassica napus* BAC library and PCR screening.

The rapeseed BAC library we used to recover genomic clones carrying target genes sequences was generated within Genoplante program on “DarmorBZH” genotype. It consists of 192 microtitreplates of 384 wells each, representing 12X of genome coverage (BAC library maintained in CNRGV, http://cnrgv.toulouse.inra.fr/library/genomic_resource/Bna-B-DarmorBZH). In order to screen the rapeseed BAC library by polymerase chain reaction (PCR), BAC clones from each of the 384-well plates were pooled and 192 pool plate DNAs were prepared using standard specification of a whole genome amplification enzyme commercial kit (Genomiphi.v2 GE Healthcare kit). Only 232 amplification reactions are needed to recover a specific clone coordinate from the initial library (192 plate pools, then 16 row and 24 column pools and 24 individual columns per positive plate).

1:200 diluted Pool plate DNAs were screened by real time PCR using gene-specific primers corresponding to the genes described in the Table S20. Real-time PCR reactions were performed on the LightCycler480 with the DNA SYBR Green I Master kit (Roche Applied Science) in a final reaction volume of 5 µl containing 1 µl of diluted BAC DNA pools, 2.5 µl of premix SYBR2X, 0.1 µl of each primer (10µM), and 1.3 µl of H2O. PCR conditions were specific to each primer couple.

b) High Density Filter production and hybridization.

Complementary to the pools, we prepared high density macroarrays for the whole Bna-B-DarmorBZH BAC library. The 192 plates were spotted on 2 nylon filters using a 6x6 pattern. For the wheat, the subset corresponding to the large fragment insert sizes from the T. aestivum cv. Chinese Spring BAC library (Allouis et al. 2003; BAC library maintained in CNRGV http://cnrgv.toulouse.inra.fr/en/library/genomic_resource/Tae-B-Chinese spring) was spotted on 7 membranes using a 7x7 pattern. These macroarrays represent 1008 plates and a 4X genome coverage. Spotting and hybridisation were performed as described in Gonthier et al. (2010).

Gene expression

For wheat we focussed on the 19 gene families for which we had confirmed the number of copies and their respective full length sequences by BAC library and PCR screening. We took advantage of available RNA-seq data to determine the contributions to the transcriptome from the A, B and D wheat genomes. For oilseed rape, for which RNA-seq data were not available we focussed on a subset of 7 meiotic recombination genes determining the contributions to the transcriptome from the A and C genomes by pyrosequencing.

Gene expression in wheat

For wheat we focussed on the 13 genes for which we had confirmed the number of copies and their respective full length sequences by BAC library and PCR screening. We took advantage of available RNA-Seq data to determine the contributions to the transcriptome from the A, B and D wheat genomes.

Total RNAs were extracted from four different stages of developing anthers (Latent/Leptotene; Zygotene/Pachytene; Diplotene/Diakinesis; Metaphase I) of the hexaploid wheat cv Chinese Spring and from 10 mg of anthers (30-50 anthers depending on the stage) using the Macherey-Nagel Nucleospin RNA-XS kit according to manufacturer instructions. RNA-Seq non-oriented libraries were constructed in duplicates using the TruSeq kit (Illumina). The eight libraries were sequenced (GATC, Germany) on two lanes (four sample per lane) of HiSeq2000 (Illumina) with paired-end sequence (500 bp) in 2x100 bp which generated 40 to 50 millions of pairs of reads per sample.

For alignment of the reads, we used the sequences of the chromosome arms produced at TGAC (UK) and available through the International Wheat Genome Sequencing Consortium (IWGSC; http://www.wheatgenome.org/). All the reads from the RNA-Seq libraries were mapped on the scaffolds representing the gene models produced from the assembly of the reads. We used TopHat2 v2.0.8 (http://tophat.cbcb.umd.edu/; Trapnell et al. 2009) with the default parameters except that we tolerated no mismatches or splice-mismatches. PCR duplicates that are generated during library construction were removed from the analysis using SAMTOOLS (rmdup option). Transcripts reconstruction and expression determination (FPKM; Mortazavi et al. 2008) were analysed with Cufflinks v2.0.2 (http://cufflinks.cbcb.umd.edu/). The genes were previously annotated and mapping of the reads was done on exons only which allowed discrimination of

expression of the homoeologous copies. Only predictions with an fpkm > 0 were considered as expressed.

RNA-seq data have been deposited in the European Bioinformatics Institute (EBI) database. Identifier: E-MTAB-2114.

Gene expression in *B. napus*

Brassica napus var. Darmor pollen mother cells (PMCs) were isolated by dissection from immature anthers and comprised cells at meiotic stages ranging from pre-meiotic interphase to meiotic prophase I. RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) with on-column DNA digestion according to manufacturer's instructions, and reverse transcribed using the BD SMART mRNA amplification kit (BD Biosciences CLONTECH), according to manufacturer's instructions.

Pyrosequencing reactions were designed with primers flanking homeologue specific SNPs in the coding sequence of meiotic genes to assess the relative contribution of each homeologue to the population of mRNA (Wittkopp et al. 2004). Pyrosequencing was performed on meiotic cDNA, and on genomic DNA as a control to normalize the ratios against possible pyrosequencing biases. Pyrosequencing primers are given in Table S24.

Supplementary References

Allouis S, Moore G, Bellec A, et al. 2003. Construction and characterisation of a hexaploid wheat (Triticum aestivum L.) BAC library from the reference germplasm “Chinese Spring.” Cereal Res. Commun. 31:331–338.

Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K. 2012. Dissecting plant genomes with the PLAZA comparative genomics platform. Plant Physiol. 158:590–600.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25:288–289.

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol. 11:136.

Cheng F, Wu J, Fang L, Wang X. 2012. Syntenic gene analysis between Brassica rapa and other Brassicaceae species. Front. Plant Sci. 3:198.

Coate JE, Schlueter J a, Whaley AM, Doyle JJ. 2011. Comparative evolution of photosynthetic genes in response to polyploid and nonpolyploid duplication. Plant Physiol. 155:2081–2095.

D’Hont A, Denoeud F, Aury J-M, et al. 2012. The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature 488:213–217.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Meth 9:772.

Dereeper a, Guignon V, Blanc G, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465–9.

Dereeper A, Audic S, Claverie J-M, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol. Biol. 10:8.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Gonthier L, Bellec A, Blassiau C, et al. 2010. Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (Cichorium intybus L., Asteraceae). BMC Res. Notes 3:225.

Goodstein DM, Shu S, Howson R, et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–86.

Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. 52:696–704.

Hamant O, Golubovskaya I, Meeley R, Fiume E, Timofejeva L, Schleiffer A, Nasmyth K, Cande WZ. 2005. A REC8-Dependent Plant Shugoshin Is Required for Maintenance of Centromeric Cohesion during Meiosis and Has No Mitotic Functions. Curr. Biol. 15:948–954.

Larkin M a, Blackshields G, Brown NP, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Lyons E, Freeling M. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 53:661–673.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. :1–8.

Murat F, Xu J-H, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J. 2010. Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res 20:1545–1557.

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. Genome Res. 13:103–107.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–12.

Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC. 2012. Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics 190:1563–1574.

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Wang K, Wang Z, Li F, et al. 2012. The draft genome of a diploid cotton Gossypium raimondii. Nat. Genet. 44:1098–1103.

Wang M, Tang D, Wang K, Shen Y, Qin B, Miao C, Li M, Cheng Z. 2011. OsSGO1 maintains synaptonemal complex stabilization in addition to protecting centromeric cohesion during rice meiosis. Plant J. 67:583–594.

Wang XX, Wang HHH, Wang JJJ, et al. 2011. The genome of the mesopolyploid crop species Brassica rapa. Nat. Genet. 43:1035–1039.

Wang Y, Tang H, Debarry JD, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. Nature 430:85–88.

Xu X, Pan S, Cheng S, et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475:189–195.

Yang Z, Nielsen R. 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol. Biol. Evol. 17 :32–43.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.