
Strong constitutive expression divergence among strains but no evidence of differential expression associated with sexual reproduction in *Alexandrium minutum*

Seveno Julie ^{1,2}, Even Yasmine ³, Le Gac Mickael ^{1,*}

¹ Ifremer, DYNECO PELAGOS, 29280 Plouzané, France

² Laboratoire Mer Molécule Santé, Le Mans Université, 72000 Le Mans, France

³ Laboratoire des Sciences de l'Environnement Marin, UMR 6539 CNRS/UBO – Institut Universitaire Européen de la Mer, Technopôle Brest-Iroise, 29280 Plouzané, France

* Corresponding author : Mickael Le Gac, email address : mickael.le.gac@ifremer.fr

Abstract :

Sexual reproduction remains poorly characterized in dinoflagellates. This is especially the case at the molecular level. Here crossing experiments were performed among strains of the toxic dinoflagellate *Alexandrium minutum* belonging to two genetically divergent groups. Gene expression was compared between sexually compatible and incompatible crosses at the time of gamete fusion and resting cyst (~zygote) formation. Not a single transcript was identified as differentially expressed between compatible and incompatible crosses at these two crucial time points of the dinoflagellate life cycle. However, several thousands of transcripts displayed constitutive expression differences between strains. This was especially the case between the strains belonging to the genetically divergent groups. A few hundreds of transcripts were also identified as differentially expressed between strains belonging to opposite mating types. Some of these transcripts displayed homology with the SxtA protein, known to be involved in saxitoxin production in cyanobacteria, as well as with proteins potentially involved in mating in fungi.

Highlights

► Sexual compatibility among divergent *Alexandrium minutum* strains. ► No evidence of differential transcript abundance during gamete fusion and encystment. ► Strong constitutive gene expression differences among strains. ► Differential gene expression among strains of opposite mating types. ► Homology with SxtA and transcripts involved in mating in fungi.

Keywords : Sexual reproduction, *Alexandrium minutum*, Gene expression, Mating

1. Introduction

Sexual reproduction is almost universal in eukaryotes. In unicellular eukaryotes, species are either homothallic when sex occurs between gametes produced by a single cell type, or heterothallic when two sexually compatible cell types are required. In heterothallic species, sexual compatibility is determined by mating types, encoded by mating-type locus (Russo et al. 2018; Hadjivasiliou and Pomiankowski 2016). The molecular bases of sexual reproduction remained unexplored in numerous unicellular eukaryote phyla, which is especially true within dinoflagellates. This group of aquatic protists is famous for its ability to form harmful algal blooms worldwide especially because some species produce toxins that bioaccumulate in shellfish, but blooming algae also may clog animal gills or deplete oxygen (Anderson et al. 2012a). Within dinoflagellates, the genus *Alexandrium* is one of the most prominent in terms of toxic blooms, with numerous species producing several toxins, especially of the saxitoxin family (Anderson et al. 2012b). This is the case of the genus type species, *A. minutum*, producing harmful algal blooms worldwide. Like most dinoflagellates, *A. minutum* displays a haplontic life cycle, with cellular division almost exclusively occurring in haploid cells (von Dassow and Montresor 2011). Sexual reproduction occurs when haploid cells fuse, leading to the formation of diploid resting cysts (Figure 1). Sexual compatibility and determinism remain poorly characterized in dinoflagellates, with systems enabling self-fertilization (homothallism) or outcrossing (heterothallism). Homothallism has been characterized in several species, for instance in the genus *Scrippsiella* (Montresor et al. 2003) or in the species *A. taylori* (Giacobbe and Yang 1999). Several species display the ability to perform both homo- and heterothallism. This is for instance the case of *Gymnodinium catenatum* (Figueroa et al. 2010) or *Pheopolykrikos hartmannii* (Chai et al. 2020). Within the *Alexandrium* genus, several species, such as *A. minutum* and *A. catenella*, seem to be strictly heterothallic, with complex compatibility systems involving more than two mating types (Figueroa et al. 2007; Mardones et al. 2016).

In the present study, crossing experiments were performed among four strains of *A. minutum*, associated with global gene expression profiling (RNA-seq), to investigate the molecular bases of sexual reproduction within this species. For each of two genetic groups displaying evidence of independent evolution and restricted gene flow based on >450,000 SNP markers (Le Gac et al. 2016), two strains displaying opposite mating type were selected based on preliminary experiments. The experimental design is summarized on Figure 2. Pairwise crosses were performed among the four strains. Global expression profiling was performed: 1. when gamete fusions were observed in the sexually compatible crosses, i.e two days after initial contact; 2. when resting cysts were first observed in the sexually compatible crosses, i.e. 5 days after initial contact; and 3. for clonal controls. Reproductive success was quantified by counting the number of resting cysts ten days after initial contact. Following this experimental design, the following questions were asked: 1. Is there sexual compatibility among the two genetic groups? 2. What are the genes displaying differential expression at the time of gamete fusion and at the time of encystment? 3. What is the level of constitutive expression among strains and are there mating type specific expression patterns?

2. Material and methods

2.1. Experimental design

Crossing experiments and RNAseq analyses were performed using four strains. The strains correspond to clonal cultures initiated after micro-pipetting individual cells from natural samples.

Strains were grown in K medium for 10 days (corresponding to the end of the exponential phase) and diluted in their own culture filtrate (0.2 μ m) in order to obtain an initial cell density of 5,000 cells per mL.

All pairwise crosses as well as clonal controls were performed in serial triplicates (all crosses from replicate one initiated in parallel on a given day, all crosses from replicate two initiated in parallel on another day...).

All cultures were performed at 18°C, under a 12:12 L:D cycle and 80 μ mol photons m⁻² s⁻¹, in algal incubators. To quantify reproductive success, for each pairwise cross, 1 mL of each strain was mixed in 24

well plates. In order to extract RNA, for each pairwise cross, 75 mL of each strain was mixed in flasks. Preliminary crossing experiments indicated that, in case of sexual compatibility, cell fusion occurred after 2 days and the first permanent cysts were observed after 5 days. Sexual reproduction was quantified using an inverted microscope, after 10 days, by counting resting cysts remaining at the bottom of the 24 well plates after washing the plates twice with sterile sea water. RNA extraction was performed after 2 days and 5 days from the same flask (Figure 2A), and we verified that the first resting cyst appeared after 5 days. For clonal controls, RNA extraction was only performed after 2 days. There was a total of 48 RNA samples. Samples are summarized in Figure 2B and in Supplementary table 1. The strains Am231 and Am233 were isolated from the Bay of Concarneau (France) in 2010, the strain Am1106 was isolated from Cork Harbor (Ireland) in 2010, and the strain Am1249 was isolated from the Bay of Brest (France) in 2013. Strains Am231 and 233 belong to *A. minutum* genetic group B. Strains Am1106 and Am1249 belong to *A. minutum* genetic group A (Le Gac et al. 2016). Mating type determination was based on a preliminary crossing experiment assessing the global pattern of sexual compatibility among ten *A. minutum* strains which suggested the occurrence of only two mating types in the tested strains (Supplementary Table 2).

2.2. RNA extraction library preparation and sequencing

For RNA extraction, cultures were centrifuged at 4500 g for 8 min and sonicated on ice for 30 s in LBP buffer (Macherey-Nagel, Duren, Germany). Extraction was performed using NucleoSpin® RNA Plus kit (Macherey-Nagel) following the manufacturer's protocol. Extracted RNA was quantified using a Biotek Epoch spectrophotometer and the quality estimated on RNA 6000 nanochips using a Bioanalyzer (Agilent Technologies, Palo Alto, CA). Library preparation was performed using the Illumina mRNA TruSeq stranded kit starting from 0.5 µg of total RNA. Library quality was assessed on a Bioanalyzer using high-sensitivity DNA analysis chips and quantified using Kappa Library Quantification Kit. Paired-end sequencing was

performed using 2 x 150 bp cycles on Illumina HiSeq3000 at the GeT-PlaGe France Genomics sequencing platform (Toulouse, France). The 48 samples were sequenced on 2 lanes (multiplex 24).

2.3. Bioinformatic analyses

Prior to read mapping, raw reads were initially characterized with FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>) and Trimmomatic (V. 0.33; Bolger et al., 2014) was used to trim ambiguous, low quality reads and sequencing adapters with parameters ILLUMINACLIP:Adapters.fasta:2:30:10, LEADING:3, TRAILING:3, MAXINFO:80:0.8 and MINLEN:70. These parameters enable: 1. To remove Illumina adapters; 2. To remove leading and trailing low quality bases; 3. To perform an adaptive quality trim balancing read length benefits with a target read length of 80 bases and sequencing error costs with a strictness factor of 0,8 (favoring read correctness); 4. To remove reads < 70 bases long after the previous steps.

A reference transcriptome previously assembled using 18 *A. minutum* strains, including the strains used in the present study was considered (Le Gac et al., 2016). This 153,222 transcripts reference, for a total of 117,601,765 bp, was annotated using UNIPROT/Swissprot (<https://www.uniprot.org/>) database and classified in Gene Ontology categories (GO; <http://geneontology.org/>) as described in Le Gac et al. (2016). Additional homology analyses were performed for transcripts displaying differential expression: first analysis: between strains displaying opposite mating types (as identified based on global compatibility patterns following preliminary crossing assays, Supplementary Table 2) and second analysis: between strains displaying opposite mating types and giving rise to the formation of resting cysts (named compatible strains hereafter; see results and supplementary Table 1). For these two sets of transcripts, an additional blastx analysis was performed against the nr ncbi database and against a subset of the nr database only containing the sequences with “mating” or “Mating” in the sequence description section. The reference transcriptome from Le Gac et al. (2016) was assembled using strains reproducing asexually

at the time of RNA extraction. A second reference transcriptome was assembled de novo using Trinity (Haas et al. 2013) after pooling the reads of the 48 samples obtained in the present study. This was done to ensure that transcripts potentially specifically expressed during sexual reproduction are considered in the differential expression analysis. From read mapping to differential expression analysis (see below), all the analyses were performed using the two reference transcriptomes. As results were similar with both reference transcriptomes, only the results obtained using the reference transcriptome from Le Gac et al. (2016) are presented. Trimmed reads were aligned to the reference transcriptome using BWA-MEM (Burrows-Wheeler Alignment Tool; Li and Durbin, 2009). Only reads with a mapping score >10 were retained. Pairs for which the two reads did not map concordantly on the same transcript were removed. Samtools was used to sort, index, and obtain raw read count table from bam files (Li et al. 2009).

2.4. Differential expression analyses

Independent analyses were performed after removing transcripts with a total read count equal to zero (149,144 transcripts considered) and transcripts with an average read count per sample lower than 10 (140,474 transcripts considered). As the two analyses gave similar results, results are only given for the second one. The raw read count table was transformed using the Variance Stabilizing Transformation (vst), and the gene expression variability across samples explored using a Principal Component Analysis (PCA). Differential expression (DE) analyses were performed using negative binomial models on raw read counts using Wald tests as implemented in DESeq2 (Love et al. 2014) and using linear models on voom normalized read counts as implemented in limma (Ritchie et al. 2015). Transcripts were considered as differentially expressed after false discovery rate (FDR) adjustment when q-values were < 0.1 using both approaches. Five types of DE analyses were performed. They all involved a simple statistical design with a single factor and two groups. More complex statistical designs involving multiple factors and interaction terms were also explored but did not lead to robust results and were therefore discarded (data not shown). The first

type of analyses was performed to identify transcripts DE when two strains are interacting in the same culture. For each pair of strains, at day 2, the expression levels quantified when the two clones have been in the same culture for 2 days and when the two clones have been in two separate cultures for two days were compared. This later condition was obtained by adding, for each pair of strains and each replicate, the read count table of the clonal controls after correcting for library size differences (contact tests; Figure 2 C1; Supplementary table 3). The second type of analyses was performed to identify, for each pair of clones, the transcripts DE between day two and five (day 2 vs day 5 tests; Figure 2 C2; Supplementary table 3). Third and fourth, expression was compared between sexually compatible and incompatible crosses at day 2 (compatibility day 2 tests ; Figure 2 C3 ; Supplementary table 3) and at day 5 (compatibility day 5 tests ; Figure 2 C4 ; Supplementary table 3). For these third and fourth analyses, the read count tables of the replicates for a given type of cross were summed and the different types of compatible crosses were considered as replicates. Based on the resting cyst quantification crosses Am231 * Am233, Am1106 * Am1249, and Am233 * Am1106 were considered as compatible and crosses Am231 * Am1249, Am233 * Am1249, and Am231 * Am1106 as incompatible (see Results). Fifth, expression was compared between pairs of strains when cultivated separately (strain tests; Figure 2 C5; Supplementary table 3). Overlap between sets of DE transcripts was analyzed using upsetR R package (Conway et al. 2017).

3. Results

3.1. Quantifying reproductive success

Crosses were performed between four strains of *A. minutum* and the number of resting cysts was quantified after ten days of contact. Our results indicated that crosses Am231 * Am233, Am1106 * Am1249, and Am233 * Am1106 produced resting cysts and were sexually compatible, while crosses Am231 * Am1249, Am233 * Am1249, and Am231 * Am1106 did not produce resting cysts (Figure 3). For

sexually compatible crosses, gamete fusion was observed at day 2 and the first resting cysts were observed at day 5.

3.2. Differential expression during sexual reproduction

Gene expression was quantified two days and five days after initial contact between strains as well as two days after inoculation for clonal controls. Gene expression variability was explored using a PCA analysis, (Figure 4).

Four types of DE analyses were performed to identify genes DE during sexual reproduction and more precisely during gamete fusion and the formation of resting cysts. The first analysis was performed to identify DE linked to the actual interaction between strains. For each pairwise cross, expression quantified after two days of contact was compared to the expression quantified in the clonal controls. As indicated Figure 5, not a single transcript was identified as differentially expressed. It was the case for the sexually compatible (Am231 * Am233, Am1106 * Am1249, and Am233 * Am1106) as well as sexually incompatible (Am231 * Am1249, Am233 * Am1249, and Am231 * Am1106) crosses. The second type of DE analysis was performed to identify transcripts DE between two and five days. Here again not a single transcript was identified as differentially expressed (Figure 6). The third and fourth types of DE analyses were performed to identify transcripts DE during gamete fusion and preceding encystment. Here again, not a single transcript was identified as differentially expressed (Figure 7).

3.3. Constitutive differential expression between strains

A fifth type of DE analysis was performed to identify transcripts DE between strains, by comparing the expression of the clonal controls. A strong DE between strains was observed, with between 6,458 and 19,616 transcripts identified as DE (Figure 8). The overlap between sets of transcripts DE in pairwise fashion is indicated on Figure 9. A total of 2389 transcripts were systematically DE between strains belonging to genetic groups A and B (Figure 9, orange barplot). Strain specific DE was identified with 1335, 1359, 1779 and 1785 transcripts displaying systematic DE in Am233, Am231, Am1106 and Am1249,

respectively (Figure 9, purple barplots). Interestingly, regarding mating type differentiation, 298 transcripts were found systematically DE between strains displaying opposite mating types (Am231 and Am1106 corresponding to mating type + and Am233 and Am1249 corresponding to mating type -; Figure 9, red barplot). Among these transcripts 183 were more expressed in mating type + strains and 115 were more expressed in mating type - strains. A total of 41 transcripts displayed homology with proteins belonging to the manually annotated UniProt/SwissProt database. Following additional blastx homology search against the nr ncbi database, two additional transcripts were identified as displaying homology with a protein known to be involved in saxitoxin production (*SxtA*). In addition, seven transcripts displayed homology with proteins potentially involved in mating in other organisms (Table 1, Supplementary Table4). Finally, there were only 43 transcripts systematically DE between strains displaying opposite mating types and which crosses lead to the formation of resting cysts (Am231*Am233, Am1106*Am1249, Am233*Am1106; not shown on Figure 9). Among these transcripts, 27 were more expressed in mating type + strains and 16 mating type - strains. A total of 6 transcripts were annotated (Table 2, Supplementary Table5).

4. Discussion

In the present study, crossing experiments were performed among four strains belonging to two genetic groups of the toxic dinoflagellate *A. minutum*.

The first main result of the study was the observation of sexual reproduction among strains belonging to two genetic groups. In a previous study (Le Gac et al. 2016), two divergent genetic groups were identified within *A. minutum*, and population genomic simulations suggested a restricted gene flow among them. Here, the ability to obtain resting cysts from crosses involving strains from the two genetic groups shows that they recognize each other as potential sexual partners and display sexual compatibility enabling the formation of resting cysts (~zygotes). However, this result should not be taken as an evidence for an

absence of reproductive isolation among the two groups, as reproductive isolation may involve postzygotic incompatibilities (inability to produce viable or fertile offspring) or environmental based reproductive isolation (that may for instance result from differences in phenology or geographic repartition; Coyne and Orr 2004). One of the two inter-genetic crosses led to the formation of resting cysts while the other did not seem to produce resting cyst. At first sight this result may suggest an asymmetric reproductive isolation (compatible A/- * B/+ crosses vs incompatible A/+ * B/- crosses). However, *A. minutum* displays a complex compatibility system involving more than two mating types (Figuroa et al. 2007). As a result, the observed asymmetry may be incidental and investigating putative asymmetric reproductive isolation would require performing crosses among numerous strains.

The second notable result is the total absence of differential expression during sexual reproduction at the time of gamete fusion as well as at the time of encystment. Dinoflagellates display numerous peculiar genomic characteristics among which the presence of spliced-leader (SL) sequences at the 5' end of most mature mRNA, suggesting that transcription in dinoflagellates may be strongly influenced by SL-trans-splicing mechanisms (Zhang et al. 2007). As indicated by previous studies, dinoflagellates polyadenylated mRNA levels are indeed rather stable with very few genes displaying differential expression across conditions (Roy et al. 2018). A possible explanation for the absence of observed differential expression could be that regulation occurs at the translation level and neither during transcription nor during mRNA maturation leading to polyadenylated mRNA. An alternative explanation would be that at the time of RNA extraction only a very small proportion of cells is actually involved in sexual reproduction, potentially resulting in the sexual reproduction specific expression pattern being too diluted to be detected at the population scale. Indeed, neglecting cell division and mortality, in the most successful crosses, about 2/5 of the initial cells ended up involved in sexual reproduction leading to the formation of a resting cyst (2500 cells.ml⁻¹ of each partner were initially mixed and about 1000 resting cysts.ml⁻¹ were obtained), but the proportion of cells actually involved in sexual reproduction at the time of RNA extraction remains

unknown. However, the presence of a sexually compatible partner could have triggered gene expression responses even in cells that are not involved in gamete fusion or encystment at the time of RNA extraction. As a matter of comparison, similar experimental designs performed in diatoms led to the identification of more than a thousand differentially expressed genes during sexual reproduction (Basu et al. 2017; Ferrante et al. 2019). A last element that could explain why differential expression during sexual reproduction in *A. minutum* was not identified may be linked to the strong constitutive expression among strains (see below). Such massive strain specific expression could mask moderate expression responses. To test this hypothesis, allele specific gene expression analyses were performed based on variable SNPs covering 21,000 transcripts (data not shown). Even using this approach, enabling the independent gene expression monitoring of each strain in a given pairwise cross, no differentially expressed transcript was identified. The last and probably most notable result of the study was the identification of massive strain specific differential expression both in terms of number of transcripts and fold change. Differential expression between strains belonging to the two *A. minutum* genetic groups was identified previously (Le Gac et al. 2016) and was confirmed in the present study with between 7 and 13% of transcripts displaying differential expression. Within the genetic groups, strain specific differential expression affected 4-6% of the transcripts, which is totally in line with a previous study in *A. minutum* (4% of differentially expressed transcripts between strains; Yang et al. 2010) and *A. fundyense* (5%; Wohlrab et al. 2016). The molecular mechanism underlying the extreme differential expression pattern observed among strains, as well as the functional consequences of such differences remain to be explored.

A few hundreds of transcripts were differentially expressed between the *A. minutum* strains displaying opposite mating types. The present study was not designed to specifically determine mating type specific expression patterns. This would require investigating differential expression among numerous strains displaying opposite mating types. However, some of these transcripts displayed notable homology with proteins potentially involved in mating in other organisms. Indeed, two transcripts displayed homology

with genes involved in mating through transfer of environmental signals involving G-protein receptors and MAPK signaling pathways, respectively. The first one corresponded to the beta subunit of a mating factor receptor-coupled G protein (Yang et al, 2012). In the yeast *Saccharomyces cerevisiae*, the G protein $\beta\gamma$ subunit is involved in the pheromone response pathway by activating signal transduction cascades involved in mating (Whiteway et al. 1995; Pryciak and Huntress, 1998). This candidate is especially interesting because, unlike what has been described in yeast, it might be sex specific in *A. minutum*. The second one, corresponded to the MAPKKK, which is part of the Mitogen-Activated Protein Kinase (MAPK) signaling pathways, one of the most important and evolutionarily conserved mechanisms to detect extracellular information in all eukaryotes. Another interesting transcript displayed homology with the JEM1p DnaJ-like chaperone located in the endoplasmic reticulum of yeast, *Saccharomyces boulardii* and *S. cerevisiae*. Several studies have shown that the $\Delta jem1$ mutant was defective in nuclear fusion during mating suggesting a key role in karyogamy during mating (Nishikawa and Endo 1997, 1998). Another candidate potentially associated with mating displayed homology with the Ste20 (Sterile) like protein kinase. This serine-threonine kinase is a member of the Ste20 related protein kinase family. This large protein family includes PAKs (p21 activated kinases) and GCKs (germinal centre kinases) and is known to be involved in mating. The Ste20-related kinases function in a MAP kinase signaling cascade and they control the mating and differentiation in the *Cryptococcus neoformans* fungus. Especially, Ste20 and Pak1 have been shown to play crucial roles in polarized morphogenesis at different steps during mating: namely cell fusion and polarity maintenance in the heterokaryotic mating filament (Nichols et al., 2004). They also have important roles in actin-mediated polarized growth. For example, deletion of Ste20 (*PAK1*) in the basidiomycete yeast *Cryptococcus neoformans* results in reduced mating-related filamentation (Wang et al., 2002). The last candidate potentially involved in mating is Pry1 (pathogen related in yeast) and together with Pry2 and Pry3, they belong to the CAP superfamily and are involved in sterol secretion in yeast cells (Darwiche et al., 2016). Among the Pry proteins only Pry3 has been demonstrated to influence mating in

yeast *S. cerevisiae* (Darwiche et al., 2018). Finally, of special interest, two transcripts displaying mating type specific expression pattern were homologous to *SxtA*, a protein known to be involved in saxitoxin production in cyanobacteria and with several known homologs in the *Alexandrium* genera (Hackett et al. 2013). Although this remains highly speculative, saxitoxin has been previously proposed to act as sex pheromone in *Alexandrium* (Wyatt and Jenkinson 1997; Cusick and Sayler 2013), and genetic differences between diverging *A. minutum* populations have been previously characterized (Le Gac et al. 2016).

5. Conclusion

The present study highlights two main results. The first one is a negative result with no evidence of differential expression associated with sexual reproduction in *A. minutum* that may directly result from dinoflagellate molecular peculiarities in terms of transcriptional regulation. Second, a strong constitutive expression difference among strains was identified, but the functional consequences of these differences remain to be explored. Finally, some transcripts with mating type specific expression patterns displayed homology with proteins known to be involved in toxin production in cyanobacteria or mating in fungi.

Acknowledgements

This work was supported by a financial support from Ifremer. We acknowledge Pascale Malestroit for crossing experiment, Julien Quéré for RNA extraction and library preparation, GetPlage for sequencing, SebiMer for bioinformatics support, and three anonymous reviewers for constructive comments.

Author contributions

MLG designed research; JS, YE and MLG analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Raw reads from the transcriptomic dataset are available (<https://doi.org/10.12770/f721cb44-3792-472f-acab-95833c401c7c>).

References

- Anderson, D.M., Alpermann, T.J., Cembella, A.D., Collos, Y., Masseret, E. and Montresor, M., 2012a. The globally distributed genus *Alexandrium*: Multifaceted roles in marine ecosystems and impacts on human health. *Harmful Algae* 14, 10-35.
- Anderson, D.M., Cembella, A.D. and Hallegraeff, G.M., 2012b. Progress in Understanding Harmful Algal Blooms: Paradigm Shifts and New Technologies for Research, Monitoring, and Management, in: Carlson, C.A. and Giovannoni, S.J. (Eds.), *Annual Review of Marine Science*, Vol 4. Annual Reviews, Palo Alto, pp. 143-176.
- Basu, S., Patil, S., Mapleson, D., Russo, M.T., Vitale, L., Fevola, C., Maumus, F., Casotti, R., Mock, T., Caccamo, M., Montresor, M., Sanges, R. and Ferrante, M.I., 2017. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist* 215, 140-156.
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Chai, Z., Hu, Z., Liu, Y. and Tang, Y. 2020. Proof of homothally of *Pheopolykrikos hartmannii* and details of cyst germination process. *J. Ocean. Limnol.* 38, 114–123.

- Conway, J.R., Lex, A. and Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938-2940.
- Coyne JA and Orr HA. 2004. *Speciation*. Sinauer Associates, Inc. Sunderland, Massachusetts USA.
- Cusick, K.D. and Sayler, G.S., 2013. An Overview on the Marine Neurotoxin, Saxitoxin: Genetics, Molecular Targets, Methods of Detection and Ecological Functions. *Marine Drugs* 11, 991-1018.
- Darwiche R., El Atab O., Cottier S., Schneiter R., 2018. The function of yeast CAP family proteins in lipid export, mating, and pathogen defense. *FEBS Letters* 8, 1304-1311.
- Darwiche R., Kelleher A., Hudspeth E.M., Schneiter R., Asojo O.A., 2016. Structural and functional characterization of the CAP domain of pathogen-related yeast 1 (Pry1) protein. *Scientific Reports* 6, 28838.
- Ferrante, M.I., Entrambasaguas, L., Johansson, M., Topel, M., Kremp, A., Montresor, M. and Godhe, A., 2019. Exploring Molecular Signs of Sex in the Marine Diatom *Skeletonema marinoi*. *Genes* 10.
- Figueroa, R.I., Garces, E. and Bravo, I., 2007. Comparative study of the life cycles of *Alexandrium tamutum* and *Alexandrium minutum* (Gonyaulacales, Dinophyceae) in culture. *Journal of Phycology* 43, 1039-1053.
- Figueroa, R.I., Rengefors, K., Bravo, I., Bens, S. 2010. From homothally to heterothally: Mating preferences and genetic variation within clones of the dinoflagellate *Gymnodinium catenatum*. *Deep-Sea Research II* 57 190–198.
- Giacobbe M. G., Yang X. 1999. The life history of *Alexandrium taylori* (Dinophyceae). *Journal of Phycology* 35, 331-338.
- Haas, B., Papanicolaou, A., Yassour, M. et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494–1512.

- Hackett, J.D., Wisecaver, J.H., Brosnahan, M.L., Kulis, D.M., Anderson, D.M., Bhattacharya, D., Plumley, F.G. and Erdner, D.L., 2013. Evolution of Saxitoxin Synthesis in Cyanobacteria and Dinoflagellates. *Molecular Biology and Evolution* 30, 70-78.
- Hadjivasiliou, Z. and Pomiankowski, A., 2016. Gamete signalling underlies the evolution of mating types and their number. *Philosophical Transactions of the Royal Society B-Biological Sciences* 371.
- Le Gac, M., Metegnier, G., Chomerat, N., Malestroit, P., Quere, J., Bouchez, O., Siano, R., Destombe, C., Guillou, L. and Chapelle, A., 2016. Evolutionary processes and cellular functions underlying divergence in *Alexandrium minutum*. *Molecular Ecology* 25, 5129-5143.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., and 1000 Genome Project Data Processing Subgroup. 2009, The Sequence alignment/map (SAM) format and SAMtools, *Bioinformatics* 25,2078-9.
- Love, M.I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Mardones, J.I., Bolch, C.J., Guzman, L., Paredes, J., Varela, D., G M. Hallegraeff, G.M. 2016. Role of resting cysts in Chilean *Alexandrium catenella* dinoflagellate blooms revisited. *Harmful Algae* 55, 238-249.
- Montesor, M., Sgrosso, S., Procaccini, G. and Kooistra, W. H. C. F. 2003. Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): evidence for cryptic species. *Phycologia* 42, 56-70.
- Nichols C.B., Fraser J.A., Heitman J. 2004. PAK kinases Ste20 and Pak1 govern cell polarity at different stages of mating in *Cryptococcus neoformans*. *Molecular Biology of the Cell* 10, 4476-4489.
- Nishikawa S., Endo T., 1997. The yeast JEM1p is a DnaJ-like protein of the endoplasmic reticulum membrane required for nuclear fusion. *Journal of Biological Chemistry* 20, 12889-12892.
- Nishikawa S., Endo T., 1998. Reinvestigation of the functions of the hydrophobic segment of Jem1p, a yeast endoplasmic reticulum membrane protein mediating nuclear fusion. *Biochemical and Biophysical Research Communications* 3, 785-789.

- Pryciak P.M., Huntress F.A., 1998. Membrane recruitment of the kinase cascade scaffold protein Ste5 by the Gbetagamma complex underlies activation of the yeast pheromone response pathway. *Genes & Development* 17, 2684-2697.
- Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 43, e47.
- Roy, S., Jagus, R. and Morse, D., 2018. Translation and Translational Control in Dinoflagellates. *Microorganisms* 6.
- Russo, M.T., Vitale, L., Entrambasaguas, L., Anestis, K., Fattorini, N., Romano, F., Minucci, C., De Luca, P., Biffali, E., Vyverman, W., Sanges, R., Montresor, M. and Ferrante, M.I., 2018. MRP3 is a sex determining gene in the diatom *Pseudo-nitzschia multistriata*. *Nature Communications* 9.
- von Dassow, P. and Montresor, M., 2011. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *Journal of Plankton Research* 33, 3-12.
- Walker L. (1984). Life histories, dispersal, and survival in marine, planktonic dinoflagellates. In: Steidinger K. & Walker L. (eds.), *Marine Plankton Life Cycle Strategies*, CRC Press.
- Wang P., Nichols C.B., Lengeler K.B., Cardenas M.E., Cox G.M., Perfect J.R., Heitman J., 2002. Mating-type-specific and nonspecific PAK kinases play shared and divergent roles in *Cryptococcus neoformans*. *Eukaryotic Cell* 12, 257-272.
- Whiteway M.S., Wu C., Leeuw T., Clark K., Fourest-Lieuvin A., Thomas D.Y., Leberer E., 1995. Association of the yeast pheromone response G protein beta gamma subunits with the MAP kinase scaffold Ste5p. *Science* 268, 1572-1575.
- Wisecaver, J.H. and Hackett, J.D., 2011. Dinoflagellate Genome Evolution. *Annual Review of Microbiology*, Vol 65 65, 369-387.
- Wohlrab, S., Tillmann, U., Cembella, A. and John, U., 2016. Trait changes induced by species interactions in two phenotypically distinct strains of a marine dinoflagellate. *ISME Journal* 10, 2658-2668.

- Wyatt, T. and Jenkinson, I.R., 1997. Notes on Alexandrium population dynamics. *Journal of Plankton Research* 19, 551-575.
- Yang, I., John, U., Beszteri, S., Glockner, G., Krock, B., Goesmann, A. and Cembella, A.D., 2010. Comparative gene expression in toxic versus non-toxic strains of the marine dinoflagellate *Alexandrium minutum*. *Bmc Genomics* 11.
- Yang R.Y., Li H.T., Zhu H., Zhou G.P., Wang M., Wang L., 2012. Draft genome sequence of CBS 2479, the standard type strain of *Trichosporon asahii*. *Eukaryotic Cell* 11, 1415-1416.
- Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T. and Lin, S., 2007. Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* 104, 4618-4623.

Table 1: Annotated transcripts DE between strains displaying opposite mating types

Transcripts	Average expression	fold change (log ₂ (MT+/MT-))	Transcript length	Homolog
comp103401_c0_seq2	349	0,50	1503	Uncharacterized protein DKFZp434B061 (Swissprot)
comp108403_c0_seq1	243	0,95	1648	Probable serine/threonine-protein kinase DDB_G0275165 (Swissprot)
comp109153_c0_seq1	291	0,78	1587	Ubiquinone biosynthesis O-methyltransferase (Swissprot)
comp111017_c0_seq1	229	1,27	1350	Pentatricopeptide repeat-containing protein At2g15630, mitochondrial (Swissprot)
comp111238_c0_seq1	429	0,71	1483	Cathepsin E (Swissprot)
comp111675_c1_seq1	546	-0,59	1273	Sterol 3-beta-glucosyltransferase UGT80A2 (Swissprot)
comp111818_c0_seq1	67	0,72	431	Elongation factor 1-alpha (Swissprot)
comp111943_c1_seq1	318	-1,68	1873	Protein MEI2-like 5 (Swissprot)
comp112946_c0_seq4	529	-2,07	3219	Eukaryotic translation initiation factor isoform 4G-2 (Swissprot)
comp113432_c0_seq1	227	1,75	1489	Protein MEI2-like 4 (Swissprot)
comp113671_c0_seq4	112	-1,00	1185	Protein MEI2-like 2 (Swissprot)
comp116500_c0_seq1	689	-1,10	1699	Mitogen-activated protein kinase 4 (Swissprot) Mitogen-activated serine/threonine protein kinase involved in mating, putative(mating)
comp117201_c0_seq1	243	-0,77	1535	Carbohydrate-binding domain-containing protein C2E1P3.05c (Swissprot)
comp117501_c0_seq1	146	-1,07	1133	SxtA long isoform, partial (nr)
comp118887_c0_seq1	1330	-0,94	2628	related to Ste20-like protein kinase; has effect on mating (mating)
comp119307_c0_seq1	810	0,57	3713	related to Ste20-like protein kinase; has effect on mating (mating)
comp120037_c0_seq1	306	-0,80	1371	D-xylose 1-dehydrogenase (NADP(+)) 2 (Swissprot)
comp124887_c0_seq1	895	-0,64	2745	Sterol 3-beta-glucosyltransferase (Swissprot)
comp125166_c0_seq1	210	-0,82	1252	Uncharacterized protein y4mH (Swissprot)
comp125447_c0_seq1	74	0,92	1524	Ribulose biphosphate carboxylase (Swissprot)
comp125574_c0_seq2	459	0,54	1738	Protein PRY1 (Swissprot) Similar to Saccharomyces cerevisiae YJL078C PRY3 Cell wall protein with a role in mating efficiency (mating)
comp127738_c0_seq1	421	0,54	2645	Cytosine-specific methyltransferase NlaX (Swissprot)
comp127932_c1_seq2	228	-1,40	1756	SxtA short isoform precursor (nr)

comp128115_c0_seq1	175	0,57	1463	Ribosomal RNA large subunit methyltransferase I (Swissprot)
comp130909_c0_seq1	449	1,08	1881	Pentatricopeptide repeat-containing protein At5g04810, chloroplastic (Swissprot)
comp132966_c2_seq1	524	-0,49	3908	mating response MAPKKK (mating)
comp133634_c0_seq1	202	0,61	2122	mating factor receptor-coupled G protein (mating)
comp136211_c0_seq1	104	-0,61	470	Ribosomal large subunit pseudouridine synthase D (Swissprot)
comp1665_c0_seq1	69	0,92	795	Ankyrin-3 (Swissprot)
comp25626_c0_seq1	154	1,58	771	Histone H4 (Swissprot)
comp35751_c0_seq1	62	0,68	1498	Pentatricopeptide repeat-containing protein At5g25630 (Swissprot)
comp42918_c0_seq2	83	0,90	949	Phosphoenolpyruvate synthase (Swissprot)
comp54009_c0_seq1	363	0,73	1652	Heme oxygenase 1 (Swissprot)
comp54880_c0_seq1	66	-0,73	593	Uncharacterized protein ycf45 (Swissprot)
comp64328_c0_seq1	70	0,80	655	Ribulose biphosphate carboxylase (Swissprot)
comp65560_c1_seq1	147	1,80	618	MOB kinase activator-like 1 (Swissprot)
comp75076_c1_seq1	107	-1,80	532	JEM1p DnaJ-like chaperone required for nuclear membrane fusion during mating (mating)
comp75209_c0_seq1	139	-2,18	1780	SET and MYND domain-containing protein DDB_G0292454 (Swissprot)
comp82904_c0_seq1	234	0,51	1001	GDT1-like protein 3 (Swissprot)
comp83927_c1_seq1	121	1,22	819	Protein aardvark (Swissprot)
comp84842_c0_seq1	93	1,44	1193	Uncharacterized protein L662 (Swissprot)
comp92889_c0_seq1	490	0,54	1487	Omega-amidase NIT2 (Swissprot)
comp93025_c0_seq1	109	-0,60	704	Elongation factor P (Swissprot)
comp95901_c0_seq1	93	1,54	1115	Mucin-19 (Swissprot)
comp98445_c0_seq1	243	1,79	1903	Ankyrin-2 (Swissprot)
comp98462_c1_seq1	350	2,97	1300	MOB kinase activator-like 1 (Swissprot)
comp98485_c0_seq2	122	-0,75	1495	Putative pirin-like protein At3g59260 (Swissprot)
comp99456_c0_seq1	280	-0,66	1281	Glycerate dehydrogenase (Swissprot)

Table 2: Annotated transcripts DE between strains displaying opposite mating types and which crosses lead to the formation of resting cysts

Transcripts	Average expression	fold change (log ₂ (MT+/MT-))	Length	Homolog
comp104439_c0_seq1	248	-0,90	1061	Probable phytanoyl-CoA dioxygenase (Swissprot)
comp120943_c0_seq2	265	0,58	2335	Pentatricopeptide repeat-containing protein At2g41720 (Swissprot)
comp123246_c1_seq1	351	0,54	2014	Calcium-dependent protein kinase 24 (Swissprot) M96 mating-specific protein family (mating)
comp129407_c0_seq1	129	-0,78	1654	FAD-containing monooxygenase EthA (Swissprot)
comp129414_c0_seq1	127	0,62	1203	Aldose reductase B (Swissprot)
comp132997_c0_seq1	103	0,60	3288	Probable helicase A859L (Swissprot)

Figure 1: Simplified *Alexandrium minutum* life cycle (adapted from Walker, 1984). Haploid vegetative cells divide by mitosis in the water column, sexually compatible cells may fuse leading to the formation of a diploid planozygote that encysts as a resting cyst in the sediment. Following a dormancy period, excystment and meiosis lead to the formation of new haploid vegetative cells.

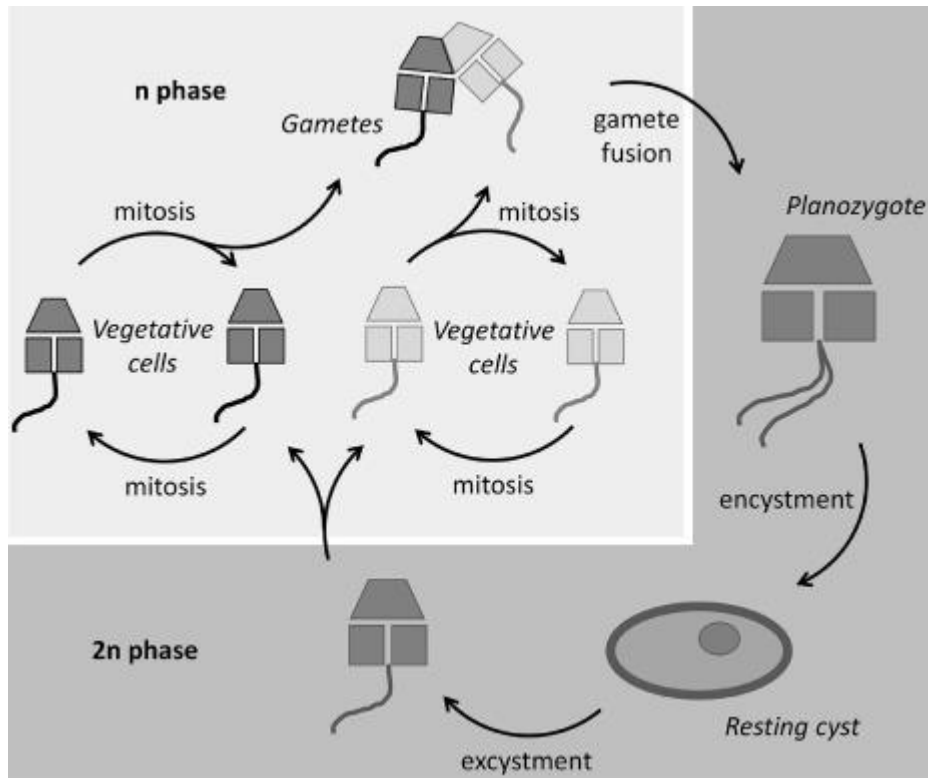


Figure 2: Experimental design. A. Analyses performed. Cells are initially mixed at day 0. RNAseq analyses are performed at day 2 (when first gamete fusions are observed) and day 5 (when first resting cysts are observed). Resting cysts are quantified at day 10 using inverted optic microscopy. B. Pairwise crosses between four strains (Am231, Am233, Am1106, Am1249) with genetic group (GG) and mating type (MT) of the strains indicated. C. Graphic summary of the five types of DE analyses performed and corresponding to the contact (1), Day 2 vs Day 5 (2), Compatibility Day 2 (3), Compatibility Day 5 (4), and strain (5) statistical analyses (see Material and Methods and Supplementary Table 3).

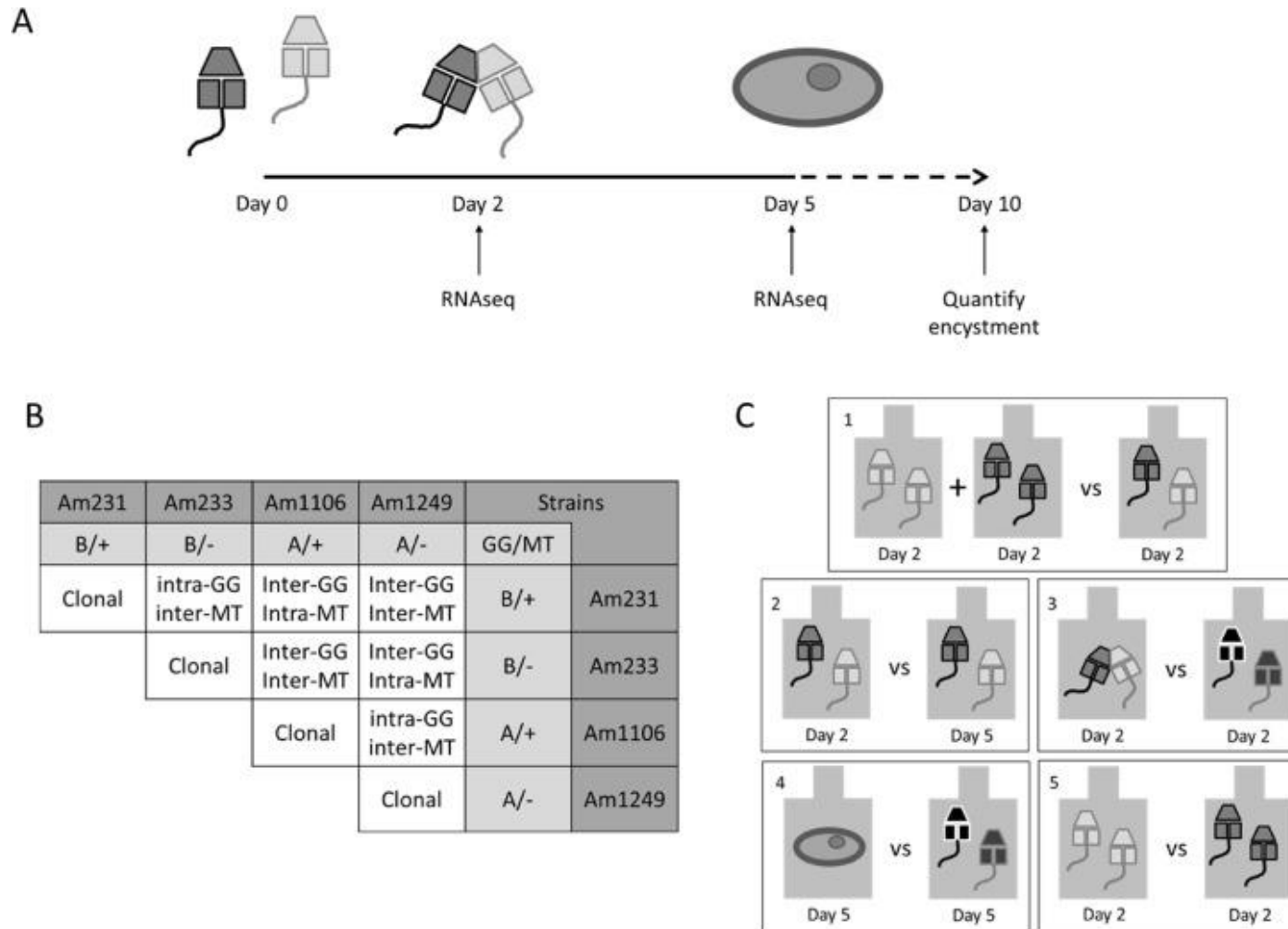


Figure 3: Resting cyst quantification performed at day 10 for the six pairwise crosses. At day 0, the initial density of each strain was 2,500 cell per ml.

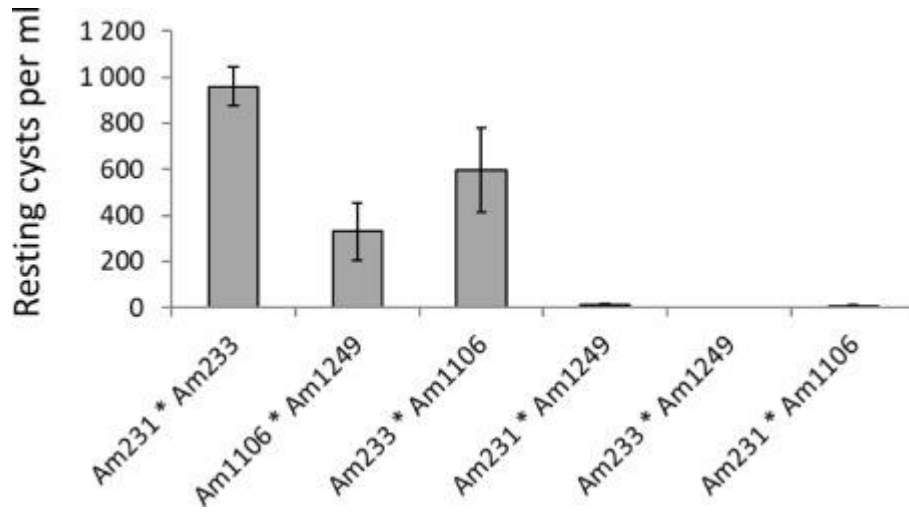


Figure 4: Overall variability of gene expression. First two axes of a PCA analysis on the entire gene expression dataset. Colors indicate the strains involved and symbols specify whether it corresponds to a clonal sample or to a pairwise cross sampled after two or five days.

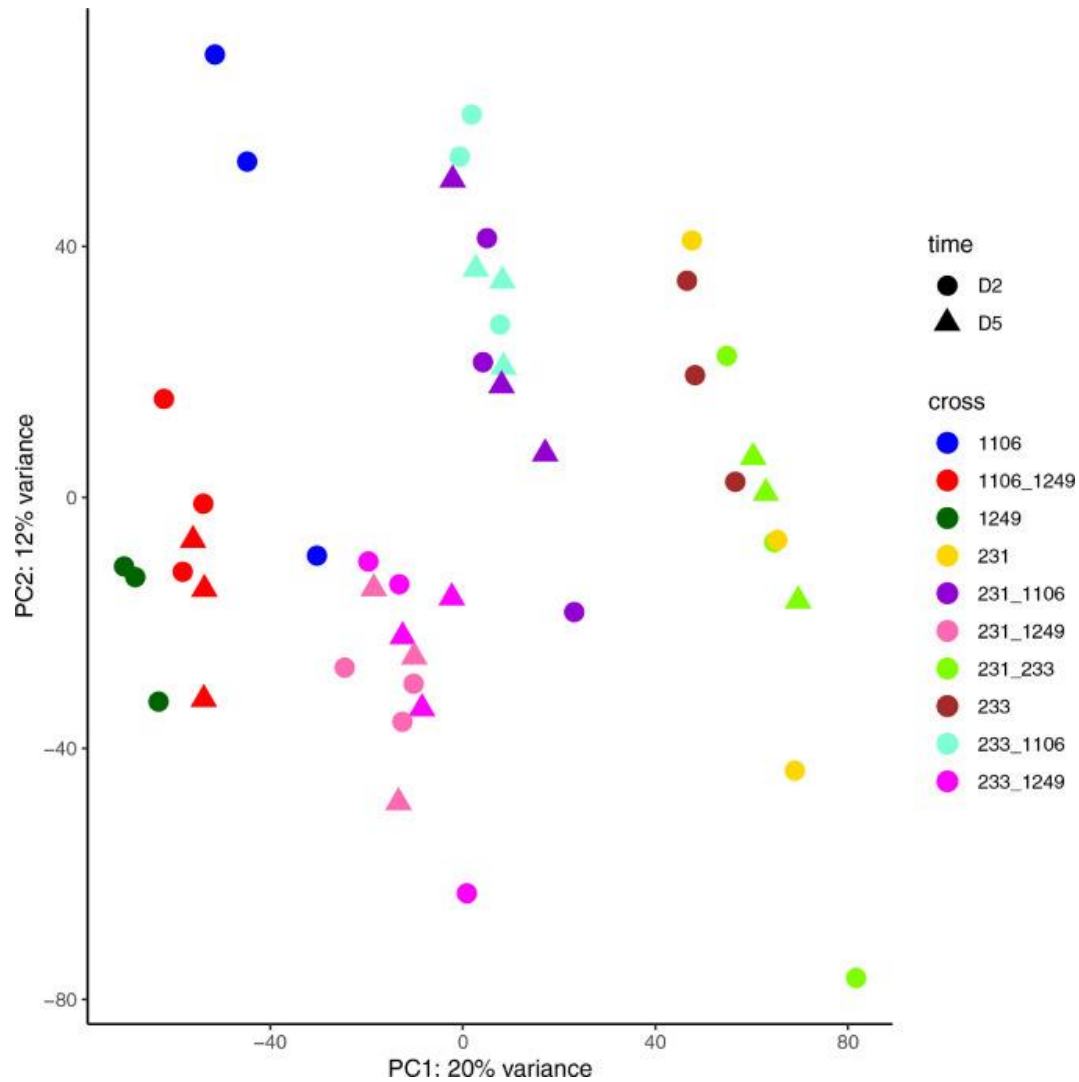


Figure 5: MA plots indicating, for each pair of strains, the difference in expression (\log_2 fold change) between clonal controls and pairwise crosses after two days. Each dot corresponds to a transcript, significant DE transcripts, if any, are in red.

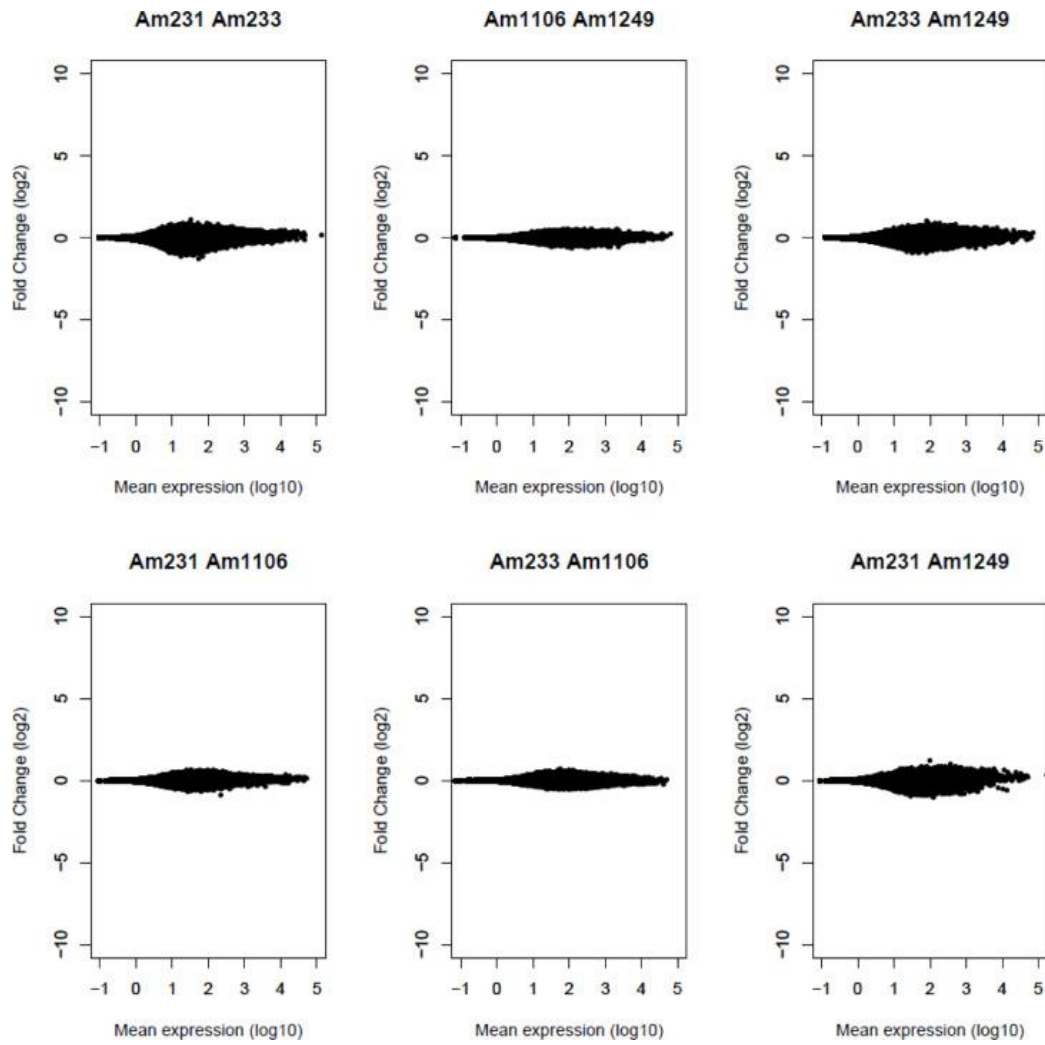


Figure 6: MA plots indicating, for each pair of strains, the difference in expression (\log_2 fold change) after two days and five days. Each dot corresponds to a transcript, significant DE transcripts are in red.

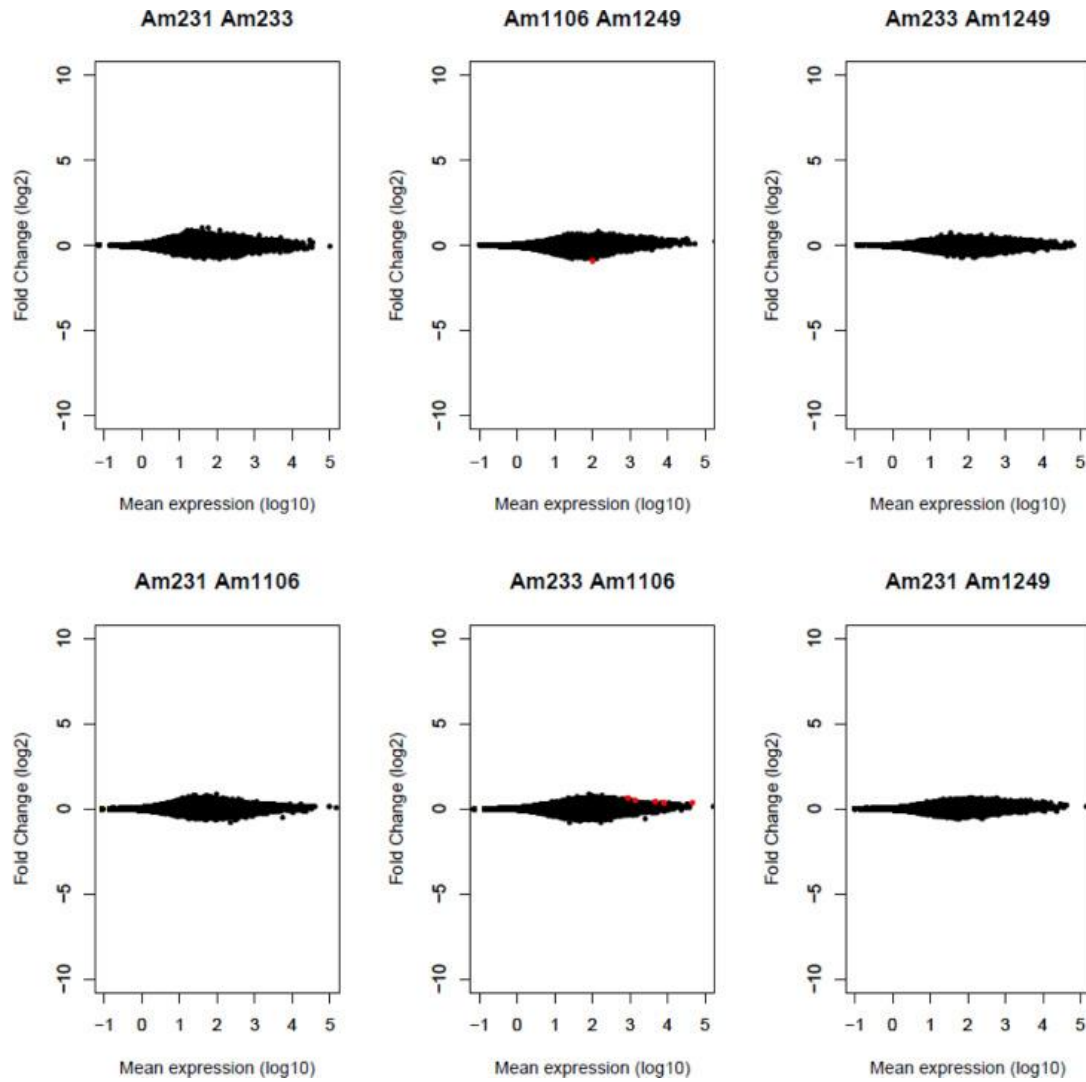


Figure 7: MA plots indicating the difference in expression (\log_2 fold change) between sexually compatible and incompatible crosses after two days and five days. Each dot corresponds to a transcript, significant DE transcripts are in red.

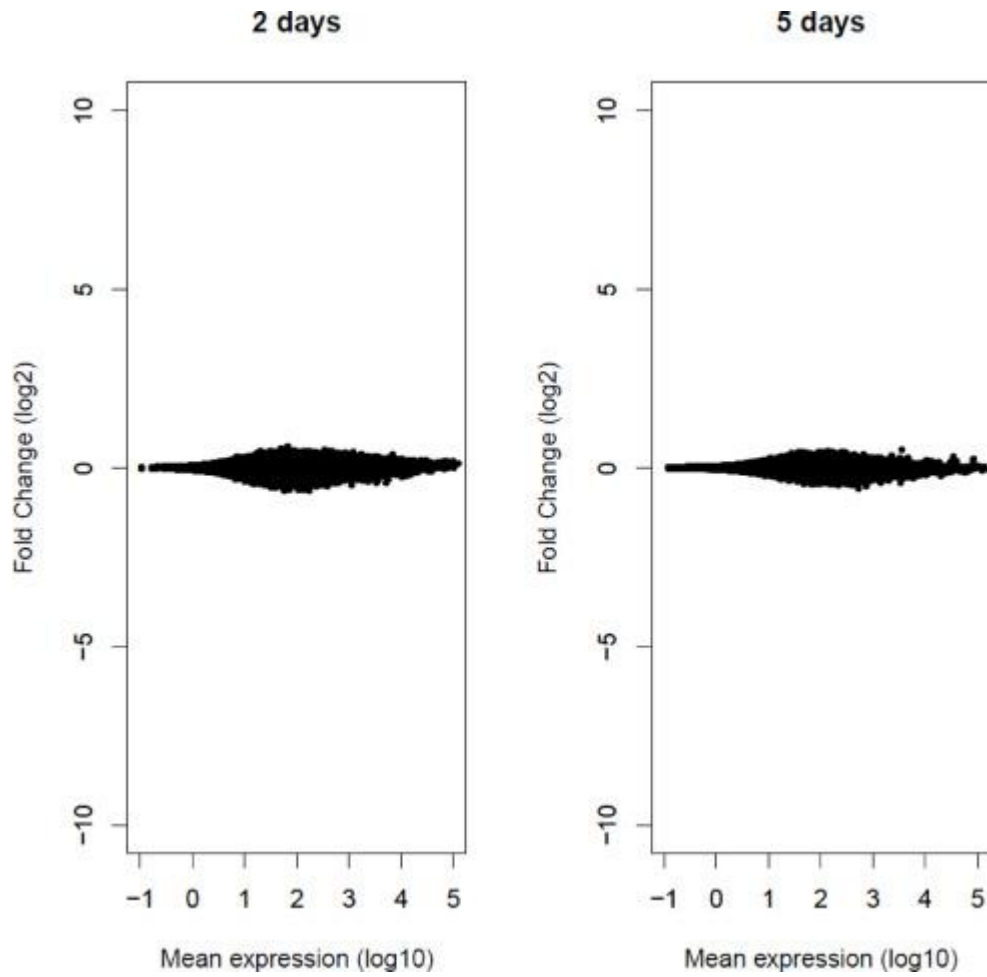


Figure 8: MA plots indicating the difference in expression (\log_2 fold change) between strains. Each dot corresponds to a transcript, significant DE transcripts are in red. The number of significant transcripts is indicated.

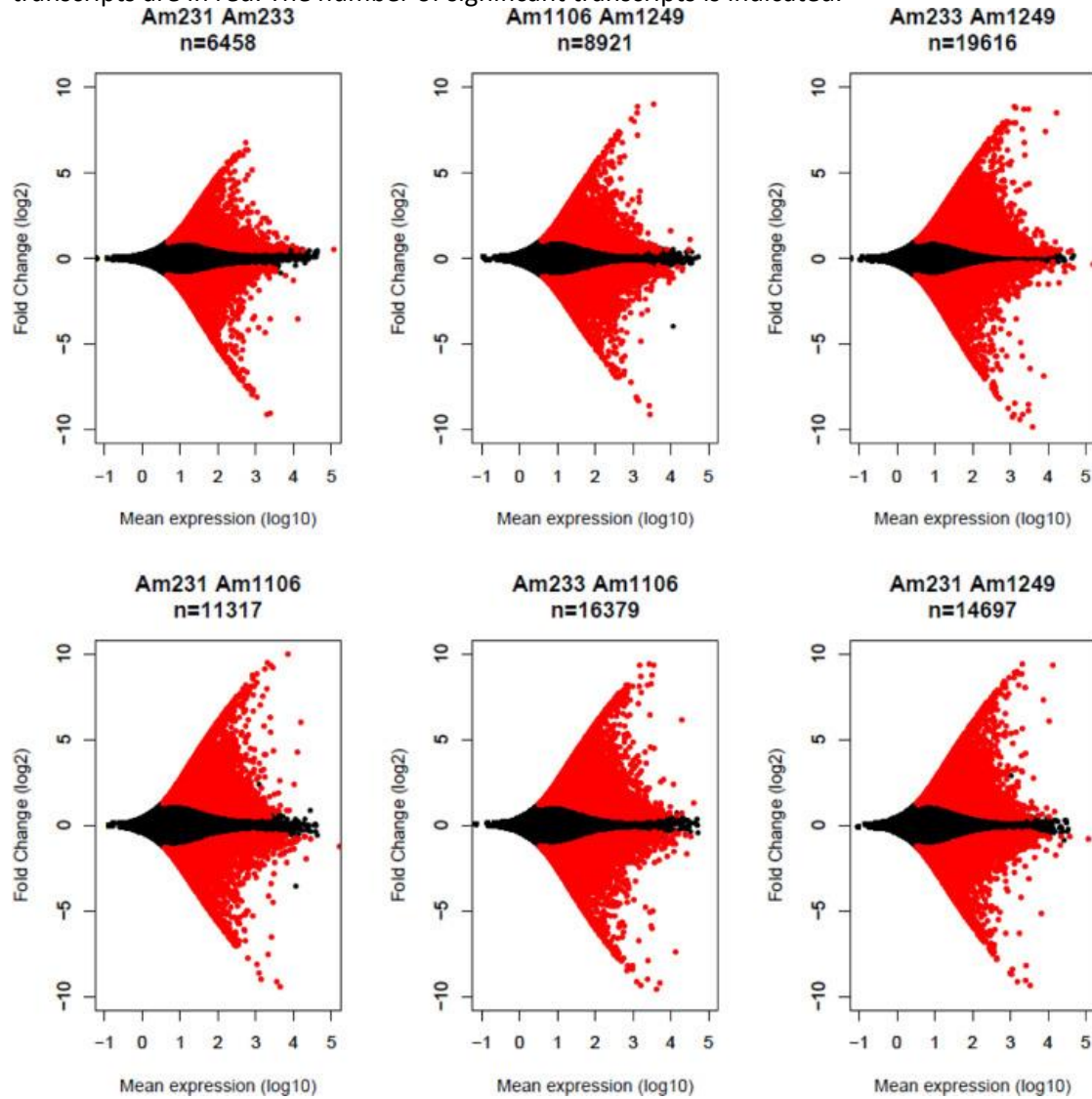


Figure 9: Overlap between sets of transcripts DE between strains. The total number of transcripts DE between pair of strains is represented on the left barplot. The intersections displaying more than 150 transcripts are represented by the bottom plot, and the number of transcripts in each intersection is shown on the top barplot. The orange barplot indicates the number of transcripts systematically DE between strains belonging to genetic group A and B. The red barplot indicates the number of transcripts systematically DE between strains belonging to mating type + and -. The purple barplots indicate the number of transcripts displaying strain specific DE.

