

Frédéric Mahé

Curriculum vitae

UMR LSTM, CIRAD
34398 Montpellier

+33 (0) 467 593 862

✉ frederic.mahe@cirad.fr



Expertise

I am an expert in bioanalysis, with a strong background in science of evolution including two years of fieldwork and data collecting experience on the Amazon River. I have specialized in the development of high-performance bioinformatics tools for targeted metagenomics and large scale ecological studies. I work on data from a wide variety of environments, with a focus on plant-associated microbial diversity in tropical and Mediterranean climates.

Education & Positions

- 2015–∞ **Researcher**, Agricultural Research for Development (CIRAD), Montpellier, France.
Diversity and impact of microbe-plant symbiosis in tropical agro-ecosystems.
- 2012–2015 **Post-Doc in ciliates genomics**, Technische Universität Kaiserslautern, Germany.
Ciliate environmental diversity and comparative genomics.
- 2010–2012 **Post-Doc in marine metagenomics**, CNRS, Station biologique de Roscoff, France.
World wide diversity of marine unicellular eukaryotes.
- 2005–2009 **PhD in plant genomics**, University of Rennes, France (December 17, 2009).
Phylogeny, transposable elements and genome size evolution in *Lupinus* L. (Fabaceae).
- 2004–2005 **Master of Science in bioinformatics**, University of Rennes, France.
Additional training in “statistics” and “machine-learning and classification”.
- 2000–2001 **Fourth year degree in population biology**, University of West Brittany, Brest, France.
Specialty “oceanography”, additional training in “marine biodiversity” and “fish biology”.
- 1999–2000 **Bachelor of Science in organism biology**, University of West Brittany, Brest, France.
Additional training in “toxicology” and “population genetics”.

Post-Doc position (2012-2015)

- title *Diversity and Genomics of Ciliates*
- supervisor Micah Dunthorn
- laboratory *Microbial Diversity* (Technische Universität, Kaiserslautern).
- abstract Biodiversity and biogeographic research in Neotropical forests is dominated by studies on plant and animal species. Our view of the functioning and conservation of these ecosystems is likewise dominated by these macro-organisms. This macro-organismic focus is primarily due to them being more familiar and readily observable to us. With the advent of massively parallel sequencing technologies, though, we are now in a better position to expand this view to include a broader microbial perspective. We normally think about Neotropical forests communities as teaming with a hyper diverse number of animals and plants, and with species assemblies changing drastically among communities because of extreme levels of endemism. By contrast, we do not know if the less familiar and not readily observable microbial eukaryotic (protist) species that inhabit these same ecosystems exhibit similar patterns of biodiversity and biogeography. This aim is to help fill in this “protistan gap” in our knowledge of these endangered ecosystems by observing soil-inhabiting ciliates in Costa Rica, Panama, and Ecuador over four years.

methodology & results These three forests will be visited once a year—during the dry season in 2012 and 2014, and wet season in 2013 and 2015. DNA & RNA from 100 soil samples will be extracted from each forest each year. Using a ciliate-specific primer approach, the hyper-variable V4 region of the small subunit ribosomal locus will be amplified, and then sequenced using Illumina MiSeq pyrosequencing. Fieldwork in Costa Rica and Panama will be in collaboration with Jordan Mayor (Smithsonian Tropical Research Institute); Gabriel Trueba (Universidad San Francisco de Quito) will collaborate in Ecuador.

I first evaluated the validity of the new Illumina MiSeq v3 as a replacement for the Roche 454 sequencing technology (Mahé *et al.*, 2014). Our results show that a higher sequencing depth and molecular diversity can be reached by the Illumina technology. Diversity profiles are comparable with the Roche 454 technology for abundance barcode sequences, but not for less abundant barcodes. Our results also allowed us to review the concept of rarity in molecular data and species counting data (Dunthorn *et al.*, 2014).

In parallel, we explored the complex genetic mechanisms of ciliate's sexual reproduction by re-analyzing published data (Chi *et al.*, 2014), and by sequencing the genomes of three ciliate species. The peculiar genomic structure of ciliates (micro- and macro-nucleus, gene-size chromosomes) poses some very interesting computational challenges, and we are still in the process of analyzing the data.

Post-Doc position (2010-2012)

title *World wide diversity of marine unicellular eukaryotes*

supervisors Colombari de Vargas & Stéphane Audic

laboratory *Evolution of plankton and paleo-oceans* (CNRS, Roscoff).

abstract Marine unicellular eukaryotes played, and are still playing, a major role in the transformations of the bio-, geo- and chimiosphere of planet Earth. Since their emergence, more than a billion years ago, the photosynthetic lineages activity precipitated large quantities of CO₂, decreased oceanic acidity, and released a large part of the O₂ present today. These dramatic modifications of the biosphere created new environments, new food chains and permitted the emergence of new modes of life (multicellularity, parasitism, or terrestrial life).

The small size (1 µm to 1 mm) and short life span of unicellular eukaryotes has complicated their study as a scientific subject. Today's advances in microscopy imaging and genetic sequencing open a new era of discoveries.

Through two ambitious international projects—BioMarKs, TARA Oceans (Karsenti *et al.*, 2011)—our objective is to explore coastal and open sea's communities of unicellular eukaryotes (including their bacterial and viral environments) and to build a collection of reference sequences covering the eukaryote's tree of life. The integration of all these data will allow us to model the dynamic system of marine communities, and to assess their capacity to respond to environmental change.

methodology & results During the BioMarKs or TARA sampling programs, large volumes of seawater are collected, filtered or fixed, at different depths and different size fractions. Organisms collected are observed, classified and quantified via an high-throughput imagery process. In a parallel metagenomic approach, DNA and RNA are extracted from samples and two regions of the rRNA small subunit are amplified with eukaryotes specific primers and pyrosequenced.

Obtained sequences are filtered, clustered, assigned to a taxonomic group, and stored in a web-accessible database. Sequences can be retrieved through SQL requests, according to a large variety of parameters, and various statistics are proposed to the end-user.

This method relies heavily on the quality and completeness of the reference database. Therefore, group-specialists participating in the project are working together in order to review and improve a manually curated database maintained at Roscoff. Local resources such as the Roscoff's cultures collection and a program of sampling, isolation, and sequencing from single cells permit us to largely improve the coverage of some poorly studied branches of the tree of life. The resulting database has been published and will be regularly updated (Guillou *et al.*, 2012).

Traditional clustering algorithms are input-order dependent and impose a fixed clustering threshold. To address these problems I developed swarm, a fast and exact clustering method, producing meaningful and fine-grain OTUs, in collaboration with Torbjørn Rognes from the Oslo University (Mahé *et al.*, 2014).

Finally, based on a genome and transcriptome sequencing program targeting 20 to 30 cultured strains of marine unicellular eukaryotes, we will build a reference database of genome sequences that will facilitate the taxonomic assignment of future metagenomics and metatranscriptomics studies.

The 1.25 billion barcode sequences collected have been analyzed, revealing new geographical and ecological patterns, as well as surprisingly high diversity in certain taxa. Appr. 15% of the barcode sequences cannot be linked to a known taxonomic group. Some of these unassigned barcodes are abundant and appear in several samples, and thus may pertain to unknown branches of the tree of life. These results are described in a *Science* paper (Vargas *et al.*, 2015).

PhD thesis

title	<i>Phylogeny, transposable elements and genome size evolution in Lupinus L. (Fabaceae)</i>
supervisors	Marie-Thérèse Misset and Abdel-Kader Aïnouche
laboratory	<i>Mécanismes à l'origine de la biodiversité</i> (Ecobio, Rennes).
abstract	<p>Genome size variation is an important trait of genome evolution, but this variation is not correlated to the complexity of organisms. Among angiosperms, genome size can vary more than 2,000-fold. Such spectacular variation raises new questions: what causes these variations? and how does this impact on the adaptation and speciation processes?</p> <p>In plants, studies show that adaptation and speciation processes correlate with a variation in genome size. Alongside whole genome duplications (polyploidy), chromosome abnormalities and DNA recombination, the accumulation or loss of repeated DNA is an important cause of genome size variation. Most of the repeated DNA is made of transposable elements, which can represent more than 50% of the genetic content in some plant's genomes.</p> <p>Thus, transposable elements, and more specifically retrotransposons, contribute significantly to the variations of plants' genome size. The structural and functional dynamic created by transposable elements—in a stress situation for instance—generate genome diversity, with consequences on adaptation and diversification of organisms.</p> <p>In this context, we first built a stronger phylogenetic framework allowing us to interpret the directionality of genome size variations. We then examined the role of retrotransposons in genome size variation and their implication in the adaptation and speciation processes in legumes of the genus <i>Lupinus</i> (Fabaceae). In parallel, and through our collaboration with the Institute of Plant Genetics (Poznań, Poland), we selected a BAC from the Polish <i>L. angustifolius</i> BAC library and obtained the very first genome data for a member of the Genistoid clade.</p>
methodology & results	<p>During my thesis, I used the following techniques: DNA and RNA extraction, PCR amplification, cloning, sequencing, cytometry and cytogenetics (FISH). I also developed a complete pipeline for sequences treatment: traces analysis, quality of reads (PHRED), removal of vector contamination (LUCY) and creation of contigs before further analysis such as statistics, comparison with databases (by blast) and phylogeny.</p> <p>An up-to-date literature focus has given me an expertise in the crucial step that is multiple alignment for phylogeny, and in its actual limitations and future developments: algorithmic asymmetry, global vs. local, meta-alignments and graph representation.</p> <p>For phylogenetic analysis, I created a front-end for the Unix version of PAUP, MRBAYES and PHYML, as a time saver for command-line non-savvy collaborators. This front-end performs format conversions, insertion-deletion coding, selection of the best model of evolution (JMODELTEST), Bremer & bootstrapping tests, nodes datation, and generation of ready-to-print trees. I selected preferentially softwares taking advantage of multi-core machines (multithreading or multiprocessing), and I now carefully monitor developments of new tools exploiting the power of graphics processors (<i>biomanycotes</i> library for instance).</p>

I also configured and used tools for expressed sequence tags annotation (EST2UNI, BLAST2GO); and genome sequences alignment, annotation, comparison and visualization (LAGAN, FGENESH, REPEATMASKER, ARTEMIS and CIRCOS).

The obtained phylogenetic framework improves our understanding of the evolutionary history of lupines (Mahé *et al.*, 2011a and Mahé *et al.*, 2011b), and when combined with the exploration of retrotransposons, highlights lineage-specific patterns of genome size variation. *Copia* and *gypsy* elements appear to contribute more significantly to genome size differences in Mediterranean lupines than in African lupines, suggesting different mechanisms among lupine lineages (Mahé *et al.*, submitted). The abundance of retroelements was confirmed at the local scale (*SymRK* gene region) where they represent more than 25% of a 110 kb gene-rich region in *Lupinus angustifolius* (Mahé *et al.*, in prep.).

Master thesis in bioinformatics

title	<i>Pygram: a new tool for genome sequence exploration</i>
supervisors	Jacques Nicolas (Irisa, Rennes) and Marc Le Romancer (LM2E, IUEM, Brest)
laboratories	<i>Symbiose</i> (Institut de recherche en informatique et systèmes aléatoires) and <i>Laboratoire de microbiologie des environnements extrêmes</i> (Institut universitaire européen de la mer).
abstract	The <i>Symbiose</i> team has developed an original and intuitive way to explore and compare genomes. It is based on a formal definition of repeated sequences and an innovative graphic user interface. The exploration of archaea genomes and associated viruses—provided by the LM2E team—demonstrated the efficiency of this approach by re-discovering, without any biological a priori, many genetic structures including CRISPRs, a new kind of immune system similar to eukaryots' microRNAs.
methodology & results	Trying to validate this new generalist tool, I confronted my observations with data from the literature. This allowed me to develop my own bioinformatics tools for the detection of remarkable genomic structures, such as CRISPRs for instance (see results in Durand <i>et al.</i> 2005, 2006). My collaboration with the <i>Symbiose</i> team is still active as we are developing new tools for CRISPR detection and analysis, at a very large scale (Mahé <i>et al.</i> , in prep.). Working as a molecular biologist with computer scientists gave me the opportunity to strengthen my bioinformatic skills, with a better foundation in mathematics and algorithmic.

Field experience

October 2014 and June 2015	Neotropical soil microbial diversity , Tiputini, Ecuador. Soil sampling mission in the protected area along the Tiputini river in Ecuador.
June 2010	BioMarKs sampling mission , Oslo, Norway. Marine water sampling mission in the Oslo Fjord.
October 2001 to July 2003	Institut de recherche pour le développement (IRD) , La Paz, Bolivia. I worked two years in a unit dedicated to the interactions between genomes, populations and environment of fresh-water tropical fishes. In charge of the sampling on the Amazon River, and the long-term conservation and samples treatment, I also participated in the data analysis process: phylogeny, phylogeography, genetics of population, morphometry, growth and reproduction, parasitology. The skills gathered in that team, the diversity of tasks and methods, and above all the splendid natural landscapes made these two years a very fruitful experience.

Languages

French	Native language	<i>Good writing skills.</i>
English	Excellent	<i>Speak, read and write English fluently.</i>
Spanish	Excellent	<i>Fluent speaker, I spent two years in Latin America.</i>

Computer skills

OS	GNU/Linux, BSD, MacOS X	publication	LaTeX, Scribus, ODF, Inkscape
language	Bash, Awk, R, Python, C11	web design	XHTML, CSS, RDF
science	data processing and visualization	hardware	assembling

Science Teaching & Tutoring experience

- 2013–2015 Fundamental concepts of bioinformatics (with Timo Mülhhaus, Master level, 10 h).
2008–2009 Introduction to algorithms for phylogeny (Master of Science in bioinformatics, 3 h).
Strategies for the detection of repeated elements (Master of Science in bioinformatics, 6 h).

Tutoring 2013–2014 internships were supervised by Micah Dunthorn, 2006–2008 internships were supervised by Abdel-Kader Aïnouche.

- 2016 Sohini Claverie (Master of Science in Biology, 6 months).
2014 Tobias Siemensmeyer (Master of Science in Biology, 6 months).
2013 Isabelle Trautmann (Master of Science in Biology, 6 months, best grade).
2008 Émilie Robin (Bachelor of Science in cellular biology, 3 months).
2007 Tristan Bitard-Feidel (Bachelor of Science in cellular biology, 1 month) and Dragomira Markova (Master of Science in ethology, ecology and evolution, 6 months).
2006 Anis Bessadok (Bachelor of Science in bioinformatics, 3 months) and Sébastien Guillotin (Bachelor of Science in ethology, ecology and evolution, 3 months).

Internships

- July 4–23, 2011 **Transcriptomics summer school**, NCGR, Santa Fe, New Mexico.
The Marine Microbial Eukaryote Transcriptome Sequencing Project will soon release 700 new unicellular eukaryotes. This unprecedented wealth of data should largely improve eukaryote phylogenies, metatranscriptoms taxonomic assignment, and our knowledge of marine metabolic pathways. Working with non-bioinformaticians, my role was to explore the data and start to answer biological questions. Internship granted by the Gordon & Betty Moore Foundation.
- September 27 to October 1st, 2010 **Grid computing for bioinformatics**, station biologique de Roscoff, France.
As biological data increase exponentially, both in size and complexity, bioinformatics treatments require more and more computing power and storage capacity. French research leaders in grid computing, web services and cloud computing share their experience through a series of practical courses. Internship granted by the CNRS.
- October 29 to December 19, 2007 **Jonathan Wendel's laboratory**, Iowa state university, USA.
After a presentation of my work, I got familiar with the tools and the bioinformatic methods used by Prof. Wendel's team for sequences cleaning (LUCY), BAC annotation (ARTEMIS) and statistical analysis of DNA chips (R). Internship granted by an "international program of scientific collaboration" CNRS-NSF.
- June 19–30, 2006 **2nd international postgraduate course in genomics**, universitat autònoma de Barcelona, Catalunya, Spain.
Laurent Duret — genomes and large-scale sequencing projects, comparative genomics, genome annotation, evolution of base composition, patterns and rates of molecular evolution.
Dmitri Petrov — polymorphism and divergence, genetic drift and selection, natural selection in the genome, structural evolution in the genome, evolution of genome size.

January–June, 2004 **Research internship**, team “évolution des génomes et spéciation”, UMR CNRS 6553 Ecobio, Rennes, France.

Value of chloroplastic data for inferring phylogeny of Lupinus L. (Fabaceae)

I tested and optimized a sequence treatment pipeline, from traces analysis to tree generation. This internship gave me the opportunity to gain advanced knowledge in the different strategies of phylogeny inference—distance-based, parametric (likelihood, Bayesian) and non-parametric (parsimony) methods.

Publications

- Durand P., **Mahé F.**, Valin A.-S. & Nicolas J. (2006) Browsing repeats in genomes: Pygram and an application to non-coding region analysis. *BMC Bioinformatics*, 7(477), 1–17.
- Chelaifa H., **Mahé F.** & Aïnouche, M. (2010) Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Molecular Ecology*, 19(10), 2050–2063.
- **Mahé F.**, Pascual, H., Misset M.-T. & Aïnouche A.-K. (2011a) New data and phylogenetic placement of the enigmatic Old World lupin: *Lupinus mariae-josephi* H. Pascual. *Genetic Resources and Crop Evolution*, 58(1), 101–114.
- Lesniewska K., Książkiewicz M., Nelson M. N., **Mahé F.**, Aïnouche A.-K., Wolko B. & Naganowska B. (2011) Assignment of three genetic linkage groups to three chromosomes of narrow-leaved lupin. *Journal of Heredity*, 102(2), 228–236.
- **Mahé F.**, Markova D., Misset M.-T. & Aïnouche A.-K. (2011b) Isolation, phylogeny and evolution of the *SymRK* gene in the legume genus *Lupinus* L. *Molecular Phylogenetics and Evolution*, 60(1), 49–61.
- Karsenti E., González Acinas S., Bork P., Bowler C., de Vargas C., Raes J., Sullivan M. B., Arendt D., Benzoni F., Claverie J.-M., Follows M., Jaillon O., Gorsky G., Hingamp P., Iudicone D., Kandels-Lewis S., Krzic U., Not F., Ogata H., Pesant S., Reynaud E. G., Sardet C., Sieracki M. E., Speich S., Velayoudon D., Weissenbach J., Wincker P. & the Tara Oceans Consortium*(Abergel C., Arslan D., Audic S., Aury J. M., Babic N., Beaufort, L., Bittner L., Boss E., Boutte C., Brum J., Carmichael M., Casotti R., Chambouvet A., Chang P., Chica C., Clerissi C., Colin S., Cornejo-Castillo F. M., Da Silva C., De Monte S., Decelle J., Desdevises, Y., Dimier C., Dolan J., Duhaime M., Durrieu de Madron X., d’Ortenzio F., d’Ovidio F., Ferrera I., Garczarek L., Garet-Delmas, M.-J., Gasol J. M., Grimsley N., Heilig R., Ignacio-Espinoza J., Jamet J. L., Karp-Boss L., Katinka M., Khalili H., Le Bescot N., Le Goff H., Lima-Mendez G., **Mahé F.**, Mazzocchi MG., Montresor M., Morin P., Noel, B., Pedrós-Alió C., Pelletier E., Perez Y., Picheral M., Piganeau G., Poirot O., Poulain J., Poulton N., Prejger F., Prihoda J., Probert, I., Rampal J., Reverdin G., Romac S., Romagnan J. B., Roullier, F., Rouviere C., Samson G., Santini S., Sarmiento H., Sciandra, A., Solonenko S., Stemmann L., Subirana L., Sunagawa S., Tanaka, A., Testor P., Thompson A., Tichanné-Seltzer V., Tirichine L., Toulza, E., Veluchamy A., Zingone A.) (2011) A holistic approach to marine eco-systems biology. *PLoS Biology*, 9(10), e1001177.
- Mella-Flores D., Mazard S., Humily F., Partensky F., **Mahé F.**, Marie D., Bariat L., Courties C., Ras J., Mauriac R., Jeanthon C., Bendif M., Ostrowski M., Scanlan D. J. & Garczarek L. (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences*, 8, 2785–2804.
- **Mahé F.**, Markova D., Lesniewska K., Książkiewicz M., Naganowska B., Misset M.-T., Wolko B. & Aïnouche A.-K. (2011) *Comparative analysis of the Symbiotic-RK genomic region between Lupinus and model legumes*. Proceedings of the 13th International Lupin Conference, Poznań, June 6-10, 56–61.

- Decelle J., Suzuki N., **Mahé F.**, de Vargas C. & Not F. (2012) Molecular phylogenetics and evolutionary history of planktonic Acantharia (Radiolaria). *Protist*, 163(3), 435–450.
- Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., Decelle J., del Campo J., Dolan J., Dunthorn M., Edvardsen B., Holzmann M., Kooistra W., Lara E., Lebecot N., Logares R., **Mahé F.**, Massana R., Montresor M., Morard R., Not F., Pawlowski J., Probert I., Sauvadet A.-L., Siano R., Stoeck T., Vaultot D., Zimmermann P. & Christen R. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41(D1), D597–D604.
- Decelle J., Martin P., Paborstava K., Pond D., Tarling G., **Mahé F.**, de Vargas C., Lampitt R. & Not F. (2013) Diversity, ecology and biogeochemistry of cyst-forming Acantharia (Radiolaria) in the oceans. *PLoS ONE*, 8(1), e53598.
- Berney C., Romac S., **Mahé F.**, Santini S., Siano R. & Bass D. (2013) Vampires in the oceans: predatory cercozoan amoebae in marine habitats. *ISME Journal*, 7(12), 2387–2399.
- Chi J., **Mahé F.**, Logsdon J., Loidl J. & Dunthorn M. (2014) Meiosis gene inventory of four ciliates reveals the prevalence of a synaptonemal complex-independent crossover pathway. *Molecular Biology and Evolution*, 31(3), 660–672.
- Dunthorn M., Otto J., Berger S. A., Stamatakis A., **Mahé F.**, Romac S., de Vargas C., Audic S., BioMarKs Consortium, Stock A., Kauff F. & Stoeck T. (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Molecular Biology and Evolution*, 31(4), 993–1009.
- Logares R., Audic S., Bass D., Bittner L., Boutte C., Christen R., Claverie J.-M., Decelle J., Dolan J. R., Dunthorn M., Edvardsen B., Gobet A., Kooistra W. H. C. F., **Mahé F.**, Not F., Ogata H., Pawlowski J., Pernice M. C., Romac S., Shalchian-Tabrizi K., Simon N., Stoeck T., Santini S., Siano R., Wincker P., Zingone A., Richards T., de Vargas C. & Massana R. (2014) The patterning of rare and abundant community assemblages in coastal marine-planktonic microbial eukaryotes. *Current Biology*, 24(8), 813–821.
- Dunthorn M., Stoeck T., Clamp J., Warren, A. & **Mahé F.** (2014) Ciliates and the rare biosphere: a review. *Journal of Eukaryotic Microbiology*, 61(4), 404–409.
- Decelle J., Romac S., Sasaki E., Not F. & **Mahé F.** (2014) Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. *PLoS ONE*, 9(8), e104297.
- **Mahé F.**, Rognes T., Quince C., de Vargas C. & Dunthorn M. (2014) Swarm: robust and fast clustering method for amplicon-based studies. *Peer Journal*, 2, e593.
- **Mahé F.**, Mayor J., Bunge J., Chi J., Siemensmeyer T., Stoeck T., Wahl B., Paprotka T., Filker S., & Dunthorn M. (2015) Comparing high-throughput platforms for sequencing the V4 region of SSU-rDNA in environmental microbial eukaryotic diversity surveys. *Journal of Eukaryotic Microbiology*, 62(3), 338–345.
- Filker S., Gimmler A., Dunthorn M., **Mahé F.**, & Stoeck T. (2015) Deep sequencing uncovers protistan plankton diversity in the Portuguese Ria Formosa solar saltern ponds. *Extremophiles*, 19(2), 283–295.

- de Vargas C.*, Audic S.*, Henry N.*, Decelle J.*, **Mahé F.***, Logares R., Lara E., Berney C., Le Bescot N., Probert I., Carmichael M., Poulain J., Romac S., Colin S., Aury J.-M., Bittner L., Chaffron S., Dunthorn M., Engelen S., Flegontova O., Guidi L., Horak A., Jaillon O., Lukeš J., Malviya S., Morard R., Mulot M., Scalco E., Siano R., Vincent F., Zingone A., Dimier C., Picheral M., Searson S., Kandels-Lewis S., *Tara Oceans Coordinators*, Acinas S. G., Bork P., Bowler C., Gorsky G., Grimsley N., Hingamp P., Iudicone D., Not F., Ogata H., Pesant S., Raes J., Sieracki M., Speich S., Stemman L., Sunagawa S., Weissenbach J., Wincker P., & Karsenti E. (2015) Eukaryotic plankton diversity in the sunlit global ocean. *Science*, 348(6237). (* all five authors equally contributed to this work)
- Morard R., Darling K., **Mahé F.**, Audic S., Ujiie Y., Weiner A., André A., Sears H., Wade C., Quillévéré F., Douady C., Escarguel G., de Garidel-Thoron T., Siccha M., Kucera M., & de Vargas C. (2015) PFR2: a curated database of planktonic Foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography, and evolution. *Molecular Ecology Resources*, 15(6), 1472–1485.
- Massana R., Gobet A., Audic S., Bass D., Bittner L., Boutte C., Chambouvet A., Christen R., Claverie J.-M., Decelle J., Dolan J., Dunthorn M., Edvardsen B., Forn I., Forster D., Guillou L., Jaillon O., Kooistra W., Logares R., **Mahé F.**, Not F., Ogata H., Pawlowski J., Pernice M. C., Probert I., Romac S., Richards T., Santini S., Shalchian-Tabrizi K., Siano R., Simon N., Stoeck T., Vaulot D., Zingone A. & de Vargas C. (2015) Marine protist diversity and community structure in European coastal areas as revealed by high-throughput sequencing. *Environmental Microbiology*, 17(10), 4035–4049.
- Richard T., Leonard G., **Mahé F.**, del Campo J., Romac S., Jones M., Maguire F., Dunthorn M., de Vargas C., Massana R. & Chambouvet A. (2015) Molecular diversity and distribution of marine fungi across 130 European environmental samples. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819).
- **Mahé F.**, Rognes T., Quince C., de Vargas C. & Dunthorn M. (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *Peer Journal*, 3, e1420.
- Kopylova E., Navas-Molina JA., Mercier C., Xu ZZ., **Mahé F.**, Rognes T., Caporaso JG. & Knight R. (2016) Open-source sequence clustering methods improve the state of the art. *mSystems*, 1(1), 00003-15.
- Le Bescot N., **Mahé F.**, Audic S., Dimier C., Garet M.-J., Poulain J., Wincker P., de Vargas C. & Siano R. (2016) Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environmental Microbiology*, 18(2), 609–626.
- Forster D., Dunthorn M., Stoeck T. & **Mahé F.** (2016) Comparison of three clustering approaches for detecting novel environmental microbial diversity. *Peer Journal*, 4, e1692.
- Forster D., Dunthorn M., **Mahé F.**, Dolan J., The BioMarKs Consortium, Audic S., Bass D., Bittner L., Christen R. Edvardsen B., Kooistra W., Logares R., Massana R., Pawlowski J., Zingone A., de Vargas C. & Stoeck T. (2016) Benthic protists: the under-charted majority. *FEMS Microbiology Ecology*, 92(8): fiw120.
- Rognes T., Flouri T., Nichols B., Quince C. & **Mahé F.** (2016) VSEARCH: a versatile open source tool for metagenomics. *Peer Journal*. (in press)
- **Mahé F.**, de Vargas C., Bass D., Czech L., Stamatakis A., Lara E., Singer D., Mayor J., Bunge J., Sernaker S., Siemensmeyer T., Trautmann I., Romac S., Berney C., Kozlov A., Mitchell EAD., Sepey CVW., Egge E., Lentendu G., Wirth R., Trueba G. & Dunthorn M. (2017) Soil Protists in Three Neotropical Rainforests are Hyperdiverse and Dominated by Parasites. *Nature Ecology & Evolution*. (in review)

Talks & Posters

- 2016 Lentendu G., **Mahé F.**, Dunthorn M. & Dunthorn M. *Biogeography of predatory protists in Neotropical forest soils*. Poster, Journées ouvertes de biologie, informatique et mathématiques (Jobim), Toulouse, June 28-30.
- Antoine-Lorquin A., **Mahé F.**, Dunthorn M., Belleannée C. & Dunthorn M. *Impact de la recherche d'amorces mutées sur les résultats d'analyses métagénomiques*. Poster, Journées ouvertes de biologie, informatique et mathématiques (Jobim), Toulouse, June 28-30.
- Mahé F.** *Metabarcoding noise control: bioinformatics progress & challenges*. Invited Talk, UniEuk, Paris, France, May 3.
- Mahé F.** *Metabarcoding: progress & challenges*. Invited Talk, CBGP, Montpellier, France, March 8.
- 2015 **Mahé F.** *Sequencing technologies & metagenomics*. Invited Talk, ISRA, Dakar, Senegal, November 17.
- Mahé F.** *Ecology in the era of big data: progress & challenges*. Invited Talk, TARA-Oceans, Océanopolis, Brest, France, October 13.
- Mahé F.** *Ecology in the era of big data: progress & challenges*. Invited Talk, University of Oslo, Norway, September 23.
- Mahé F.** *Environmental metagenomics with Swarm*. Invited Talk, University of Brest, France, March 26.
- 2014 **Mahé F.** *Swarm: robust and fast clustering method for amplicon-based studies*. Invited Talk, Unité Génétique, Physiologie et Systèmes d'Élevage, INRA, Toulouse, France, September 22.
- Mahé F.**, Rognes T., de Vargas C., Quince C. & Dunthorn M. *Swarm: robust and fast clustering method for amplicon-based studies*. Poster, 13th European Conference on Computational Biology, Strasbourg, France, September 7-10.
- Mahé F.** *Swarm: robust and fast clustering method for amplicon-based studies*. Invited Talk, Workshop on Recent Computational Advances in Metagenomics, Strasbourg, France, Sept. 7.
- Mahé F.** *Clustering and taxonomic assignment methods for amplicon-based environmental diversity studies*. Invited Talk, Exploring and testing eukaryotic-omics with next generation sequencing, Geneva, April 24-25.
- 2013 **Mahé F.**, Rognes T., de Vargas C., Quince C. & Dunthorn M.. *Swarm: fast & exact amplicon clustering*. Poster, 2nd Colloque de Génomique Environnementale, Rennes, November 4-6.
- Martin G., Samar M., **Mahé F.**, Coriton O., Novak P., Macas J. & Ainouche A.-K. *Genome size variation and evolutionary dynamics of the repetitive compartment in Mediterranean and African lupines (Lupinus; Fabaceae)*. Poster, 2nd Colloque de Génomique Environnementale, Rennes, November 4-6.
- Mahé F.**, Rognes T., de Vargas C. & Dunthorn M. *Swarm: fast & exact amplicon clustering*. Poster, Journées ouvertes de biologie, informatique et mathématiques (Jobim), Toulouse, July 1-4.
- Mahé F.**, Rognes T., de Vargas C. & Dunthorn M. *Transparent clustering method for cryptic molecular species of protists*. Talk, Joint meeting between the British Society for Protist Biology and the Linnean Society: Protists, other small organisms, and Next Generation Sequencing, London, April 10-12.
- 2012 **Mahé F.**, de Vargas C. & The Tara Oceans Consortium. *TARA OCEANS, exploring world-wide marine biodiversity*. Talk, Analyse bio-informatique des données de métagénomique, Lille, December 19.
- Mahé F.**, Martin G., Coriton O., Salmon A., Misset M.-T., Neumann P., Novak P., Macas J. & Ainouche A.-K. *Genome size variation and transposable element dynamics in Mediterranean and African Lupinus genomes*. Poster, 20th International Plant & Animal Genome, San Diego, January 14-18.

- 2011 **Mahé F.** *The World of CRISPRs—the immune system of bacteria and archaea*. Talk (best talk award!), 3rd Young Scientists Day, Roscoff, November 24th.
- Morard R., Audic S., Santini S., Poulain J., Le Bescot N., **Mahé F.**, Romac S., Not F., Ujjié Y., Wincker P. & de Vargas C. *Reaching the boundaries of global planktonic foraminiferal biodiversity using Next Generation Sequencing*. Poster, Modern Planktic Foraminifera and Ocean Changes, Amsterdam, September 1st.
- Decelle J., Suzuki N., **Mahé F.**, de Vargas C. & Not F. *Molecular phylogenetics and evolutionary history of planktonic Acantharia (Radiolaria)*. Poster, BioSystematics, Berlin, February 21-27.
- Mahé F.**, Markova D., Lesniewska K., Książkiewicz M., Naganowska B., Misset M.-T., Wolko B. & Ainouche A.-K. *Comparative analysis of the Symbiotic-RK genomic region between Lupinus and model legumes*. Talk, 13th International Lupin Conference, Poznań, June 6-10.
- 2010 **Mahé F.**, Audic S., Christen R., Claverie J.-M., Ogata H., Dolan J., Edvardsen B., Kooistra W., Massana R., Pawlowski J., Richards T., Stoeck T. & de Vargas C. *Exploring the biodiversity of the world largest ecosystem: BioMarKs project first results and bioinformatics challenges*. Poster, Journées ouvertes de biologie, informatique et mathématiques (Jobim), Montpellier, September 7-9.
- Audic S. & **Mahé F.** *The POSEIDON project: protistian diversity as revealed by ribosomal barcoding along the TARA OCEANS' circumnavigation*. Talk, 8th meeting of the GenOuest Bioinformatics Platform (special "NGS technics") 2010, Rennes, October 26.
- 2009 Ainouche A.-K., Affagard M., **Mahé F.**, Ainouche M. & Misset M.-T. *Molecular evidence for an allopolyploid origin of the invasive european gorse, Ulex europaeus subsp. europaeus (Fabaceae; Genisteae) — & — Chelaifa H., Mahé F., Monnier A. & Ainouche M. Cross species hybridization microarrays reveal consistent transcriptomic changes following natural interspecific hybridization and allopolyploid speciation in Spartina (Poaceae)*. Posters, International Conference on Polyploidy, Hybridization and Biodiversity, Saint-Malo, May 17-20.
- Mahé F.**, Leśniewska K., Książkiewicz M., Naganowska B., Misset M.-T., Wolko B. & Ainouche A.-K. *Legume genomics: comparative analysis of the SymRK region among representatives of different papilionoid lineages*. Poster, Evolutionary Biology 2009, Rennes, June 8-12.
- Mahé F.**, Biteau C., Robin É., Pasquet R., Catrice O., Huteau V., Brown S., Coriton O., Misset M.-T. & Ainouche A.-K. *Diversity and contribution of retrotransposons to lupines' genome size evolution*. Talk, 16th national conference on transposable elements, Le Mans, July 1-3.
- Książkiewicz M., Leśniewska K., **Mahé F.**, Naganowska B., Ainouche A.-K., Misset M.-T. & Wolko B. *Nodulation SymRK-gene region in the narrow leafed lupin genome: localisation, structure and annotation*. Poster, 4th Conference of Polish Society of Experimental Plant Biology, *Acta Biologica Cracoviensia*, 51(2), 79. Kraków, Poland. September 21-25.
- 2008 **Mahé F.**, Biteau C., Robin É., Pasquet R., Catrice O., Huteau V., Brown S., Coriton O., Misset M.-T. & Ainouche A.-K. *Diversity and differential evolutionary patterns of retrotransposons among lupins (Lupinus L.; Fabaceae) with variable genome size adapted to contrasted environmental conditions*.
- Talk, conference on "Cytogenetics and polyploidy", Angers, April 2-4.
 - Poster, 1st international conference on transposable elements, Saint-Malo, April 20-23.
 - Talk, 11th conference on "Plant biotechnology and sustainable management of stress", Rennes, June 30-July 3.
 - Talk, 30th meeting of the "biology and population genetics group", Rennes, August 25-28.
 - Poster, 7th national conference on "Genetic resources", Strasbourg, October 13-15.
- 2007 **Mahé F.**, Misset M.-T. & Ainouche A.-K. *Exploring retrotransposon diversity and evolution in Lupinus (Fabaceae) in the context of species diversification and adaptation to contrasted environmental conditions*.
- Poster, conference on "Environmental genomics: from individual genomes to genomes of complex communities", Roscoff, June 9-13.
 - Poster, 15th national conference on transposable elements, Rennes, July 3-5.

- 2006 **Mahé F.**, Sinou C., Pasquet R., Misset M.-T. & Ainouche A.-K. *Identification of Ty1-copia retrotransposon families in Lupinus by isolation and phylogenetic analysis of reverse transcriptase sequences.*
- Poster, 14th national conference on transposable elements, Clermont-Ferrand, July 3-5.
 - Poster, Botanical Society of America, Chico, California state university, July 28-August 2.

Review

Referee for the following peer reviewed publications: Acta Protozoologica, BMC Bioinformatics, BMC Research Notes, Environmental Microbiology, FEMS Microbiology Ecology, ISME Journal, Journal of Eukaryotic Microbiology, Journal of Plankton Research, Molecular Ecology, Molecular Ecology Resources, Molecular Phylogenetics & Evolution, mSystems, Nature methods, PLoS ONE, Polar Biology, Research in Microbiology and Scientific Reports.