



Historical note/Note historique

## Ten years of the Génolevures Consortium: A brief history

### *Les dix ans du consortium Génolevures : un bref historique*

Jean-Luc Souciet

for the Génolevures Consortium (GDR CNRS 2354)

CNRS UMR 7156, laboratoire « génétique moléculaire génomique microbiologie », université de Strasbourg, institut de botanique, 28, rue Goethe, 67083 Strasbourg cedex, France

## ARTICLE INFO

## Article history:

Received 19 April 2011

Accepted after revision 9 May 2011

Available online 1 July 2011

## Keywords:

Yeasts

Génolevures consortium

## Mots clés :

Levures

Consortium Génolevures

## ABSTRACT

We report the early phase of yeast comparative genomics conducted by a group of seven French CNRS laboratories: the Génolevures Consortium. This first multispecies comparison of Hemiascomycetes (now called *Saccharomycotina*) opened the way to yeast evolutionary genomics. This analysis indicates that yeasts are powerful organisms to decipher the different mechanisms acting to reshape the genome of eukaryotes during long evolution periods. This initial DNA approach reveals how biodiversity could be characterized with robust data.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

## R É S U M É

Nous présentons la contribution du consortium Génolevures, qui regroupe sept laboratoires français associés au CNRS, dans l'émergence de la génomique comparative des levures. Cette comparaison entre espèces appartenant au groupe des Hémiascomycètes est à la base de la génomique évolutive. La démarche utilisée montre le rôle déterminant que joue la génomique comparative appliquée aux levures pour identifier les mécanismes impliqués dans l'évolution des génomes d'eucaryotes. On observe également que ces données servent à caractériser la biodiversité sur des bases moléculaires robustes.

© 2011 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

### 1. Introduction

The year 2010 corresponds to the tenth anniversary of the publication by *FEBS Letters* of a special issue on Comparative Genomics [1], at this time a new emerging field in science. It is a good opportunity to retrace by a brief summary the key advances using yeast comparative genomics for evolutionary studies. The generous offer of the French "Académie des sciences" to host our consortium for a one-day meeting in the prestigious building of the Quai Conti and to publish its scientific content in the *Comptes Rendus Biologies* is acknowledged here as a highly

appreciated privilege by all members of Génolevures Consortium.

The early phases of genomics, after application of the manual methods that can be traced back to 1977, included the building of large sequencing centers such as, in 1992, the Sanger Center in UK or the Genome Center in Saint Louis, Missouri, or the Génoscope in France in 1996. Next to the human genome, for which these centers were built, complete genome sequencing started with model organisms such as the yeast *Saccharomyces cerevisiae*, the worm *Caenorhabditis elegans* and a few prokaryotes (bacteria and archaea). Given the cost and effort to obtain each genome sequence, few people at this time imagined to sequence other organisms; most were eager to initiate "post-genomic" studies as they liked to call what was indeed

E-mail address: [jlsouciet@unistra.fr](mailto:jlsouciet@unistra.fr).

largely classical analyses of gene functions. The yeast community did not escape this trend, albeit *S. cerevisiae* was the first eukaryotic genome fully sequenced [2], even if it were able to rapidly produce a large collection of single gene disruptants for systematic functional studies, a move that contributed to make *S. cerevisiae* today a reference genome for studies on other organisms.

## 2. Early phase of yeast comparative genomics

By 1997, exploring novel yeast species at the genomic level was beyond reach. The attempt made by the Institut Pasteur team [3] illustrated the power of comparative sequence analysis with *S. cerevisiae* to rapidly identify novel genes of *Kluyveromyces lactis*, but techniques were not sufficient to cover a significant share of this genome. This is when Génoscope was building sequencing power to focus on the human genome, and decided to offer a small part of its sequencing facilities to additional side projects to enlarge the variety of its expertise. A group of 7 French laboratories,<sup>1</sup> all associated to CNRS, issued a common proposal asking Génoscope for a total of ca. 50,000 Sanger reads to analyze 13 new yeast genomes scattered across the entire subphylum of budding yeasts or Hemiascomycetes (now called *Saccharomycotina*) [4]. The figure looks ridiculous by today standards but sequences were paired at the ends of 4–5 kb long inserts of yeast genomic DNA cloned into bacterial plasmids (ensuring a view on local gene synteny conservation) and were exceptionally long and high quality thanks to the LiCor machines (facilitating gene identification using simple *blast* comparisons against *S. cerevisiae*, or the meager dataset of other genomes available at this time). With this technique of paired Random Sequence Tags (RST), corresponding to merely 0.2–0.4 X genome coverage, but manually annotated from *blast* results, a total of 22,000 new genes could be unambiguously identified, the largest dataset from a single eukaryotic phylum at this time. With such data, annotation of the *S. cerevisiae* genome could be improved (50 novel genes identified), ascomycete-specific genes were distinguished from common genes, showing that they tend to diverge more rapidly in sequence during evolution, conservation of synteny could be quantified, deviation of the genetic code was identified as well as important size variation in gene families in the various hemiascomycete branches, consistent with functional properties. The existence in other yeast species of genes absent from *S. cerevisiae* pinpointed the gain and loss of genes during evolution, a phenomenon now known to be of primary importance in all organisms. All these results appeared, in 2000, in a special issue of *FEBS Letters* called “Génolevures: Genomic Exploration of the Hemiascomycetous Yeasts” [1]. The Génolevures Consortium was born.

Other laboratories subsequently used a similar RST approach with different yeast species to identify genes and regulatory elements in *Saccharomyces* yeasts, or to prove the whole-genome duplication (WGD) hypothesis in their

ancestry [5–8]. The RST approach alone, however, was not sufficient to address other questions such as, for example, the genome redundancy, the evolution of multigene families, or the reshaping of chromosomal maps. Complete genome data were needed.

## 3. The Génolevures Consortium second phase (2001–2004): the first multispecies comparisons of yeast genomes opened the way to yeast evolutionary genomics

Fortunately, in 2002, Génoscope accepted a new Génolevures proposal to fully sequence four yeast genomes as a means to explore the evolution of hemiascomycetes as broadly as possible. The decrease of synteny conservation we had observed with our first 13 yeast species as compared to *S. cerevisiae* [9], was used as a criterion to select three species to represent this phylum as broadly as possible: *K. lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica*. A fourth species, absent for our initial 13 but whose partial sequence had been determined at the Pasteur Institute, was added to this set: the human pathogen *Candida glabrata*, phylogenetically closely related to *S. cerevisiae*.

These four fully sequenced new yeast genomes, manually annotated by experts, were used together with *S. cerevisiae* for multiple comparisons out of which major evolutionary signatures within the hemiascomycete subphylum started to emerge [10]. For example: (i) *Y. lipolytica* with its ca. two-fold larger genome differs from the four other yeasts that seem to share a genome size control mechanism; (ii) compact and well-defined centromeres are detectable for *S. cerevisiae*, *C. glabrata* and *K. lactis*, but not in *Y. lipolytica* and *D. hansenii*; (iii) in *D. hansenii*, paralogous genes increase in number by the formation of more numerous tandem repeats than in other yeasts; (iv) *C. glabrata*, although resulting from the WGD event, shows the poorest level of duplication of protein coding genes and, in addition, a functionally coordinated loss of genes, suggesting a reductive evolutionary scheme.

The analysis of amino-acid sequence divergence, calculated for the entire set of orthologous proteins between the newly-sequenced yeasts and *S. cerevisiae*, revealed that the evolutionary distance separating *S. cerevisiae* from *Y. lipolytica* is larger than the one observed for the entire phylum of chordates. In addition to a total of ca. 24,200 novel genes identified, we observed a large diversity of evolutionary events affecting each analyzed branch, indicating that yeasts are powerful organisms to decipher how different and overlapping mechanisms are acting to reshape the genome of eukaryotes during long evolution periods.

## 4. Analysis of genome redundancy and conservation of synteny using the protoploid *Saccharomycetaceae* species

Numerous events sculpt the genomes in each lineage, resulting in specific or common features. One of the latter is genome redundancy, a common feature of nearly all eukaryotic genomes sequenced so far. This redundancy results from the combination of different events such as segmental duplications, retroposition, aneuploidy or the whole-genome duplication event in the ancestry of *Saccha-*

<sup>1</sup> Institut Pasteur, INRA Agro-Paris-Tech, University of Paris11, University of Lyon 1, University of Bordeaux 2, University of Strasbourg.

romyces and related yeasts [11] whose consequences have focused much attention in evolutionary studies.

The taxonomy of Hemiascomycetous yeasts (recently renamed *Saccharomycotina* subphylum by Kurtzman et al. [4]) has undergone numerous modifications from the end of the nineteenth century, mostly depending of the usage of morphological or biochemical criteria [12] and, more recently, genomic criteria prone to further changes when additional DNA sequences became available. From presently available sequences [13], three major subdivisions emerge within *Saccharomycotina* corresponding roughly to the *Dipodascaceae* family (small number of chromosomes and dispersed 5S RNA genes); the CTG group (deviation of the genetic code) and the *Saccharomycetaceae* family (characterized by highly conserved short centromeres and triplication of mating-cassettes), but additional branches need to be explored at the genomic level. The *Saccharomycetaceae* family can be further subdivided into two subgroups based on the WGD event, leaving a complex of phylogenetic branches escaping this event that we collectively called “protoploids” as we concentrated our study on them. By sequencing three new genomes of this complex: *Kluyveromyces thermotolerans*, *Saccharomyces kluyveri*, *Zygosaccharomyces rouxii*, and including in comparisons two previously sequence genomes of the same complex: *K. lactis* [10] and *E. gossypii* [7], we were able to define the common gene set of *Saccharomycetaceae* and illustrate the impact of redundancy for yeast genome evolution. The comparison between protoploid and post-WGD genomes [14] revealed similar levels of gene redundancy (1.25 on average), indicating that this intrinsic genome property is quantitatively more related to the succession of minor individual duplication events than to the WGD. The limited fluctuation in total gene number between species, and the simultaneous presence of ancient and recent paralogs in each one, support the view that active genome dynamics is the most important driving force for genome evolution.

A new method to define orthologs was proposed and used to identify synteny blocks in protoploid *Saccharomycetaceae*. Even the most distant species among the five studied, share a significant conservation of synteny, and that the most important source of novelty stands with a fast amino-acid substitution rate. From conserved synteny blocks, at least two classes of intervening genes were distinguished: those corresponding to recent formation of paralogs, and those representing traces of horizontal gene transfer. This distinction may be of general significance when studying eukaryotic species sharing high enough levels of synteny conservation.

##### 5. A database dedicated to yeast comparative genomics: <http://www.genolevures.org>

The use of a well-defined nomenclature facilitates comparative genomic analyses and data interpretation of fully sequenced genomes: it should be based on the coordinates of each genetic element along chromosomes rather than on functional similarities inferred from sequence homology and prone to evolve with time. The nomenclature proposed by Durrens and Sherman [15]

fulfils this requirement [16] and was used, at least in part, by other groups working in yeast genomics [17,18]. Its utilization by the Génolevures database (below) facilitates comparisons and the navigation procedure by attributing a single, invariant and non-ambiguous designation to each genetic element irrespective of its functional annotation.

The results and data produced by our Consortium should be easily communicated to the scientific community. A Génolevures web-site (<http://www.genolevures.org/>), funded with the help of CNRS, is offering this opportunity. Seven fully sequenced genomes of hemiascomycetous yeasts, representative of this subphylum, are now accessible with their complete annotations and with dedicated tools for comparative genomic studies: *C. glabrata*; *Z. rouxii*; *K. lactis*; *K. thermotolerans*; *D. hansenii*; *Y. lipolytica* (sequenced by Génolevures). In addition, the genomes of *S. cerevisiae*; *S. kluyveri* and *A. gossypii* genomes (sequenced by other teams) are also represented using the same criteria. This site is now well known and used by numerous scientists as indicated by the ClustrMap on the Génolevures web-site front page [16].

##### 6. A consequence of the pioneer work of Génolevures? The San Feliu de Guixols meetings: a forum to discuss the diversity of evolutionary routes taken by eukaryotic microorganisms

The rapidly growing importance of comparative genomics convinced two European Scientific Organizations, ESF and EMBO, to support in 2005 a novel international meeting organized by the Génolevures Consortium. This successful meeting, held at San Feliu de Guixols (Spain), was the place for scientists to discuss their most recent results on the genome architectures, major evolutionary events and the possible transitions at the origin of the various phyla of unicellular eukaryotes. This meeting became the first of a now well-established EMBO Conference Series (<http://cwp.embo.org/cfs09-07/>), offering scientists every odd year an ideal place to promote and stimulate collaborations and follow the extremely rapid progress of genomics of unicellular eukaryotes in the most pleasing environment of San Feliu de Guixols.

##### 7. Biodiversity and a new exploratory program: DIKARYOME

It may be obvious, but we repeatedly observed that each time we sequenced a new yeast species, its genome revealed non-expected features with important significance for our understanding of eukaryotic genome evolution. Following this experience, and to extend our genomic knowledge of hemiascomycetous yeasts as broadly as possible, we have recently sequenced two new yeast species. First, the detailed analysis of *Pichia sorbitophila* (CTG group) reveals a hybrid genome formed between distantly related parental species (ca. 15% of nucleotide divergence) and in an early phase of resolution post-hybridization. This genome unravels some of the steps leading to speciation, a situation rarely observed in a researcher's lifetime. Moreover, it suggests that interspecies hybridization is probably more common in yeasts than

previously anticipated. This analysis also suggests that in the next future, the combination of sequencing tools and comparative genomic analyses will be powerful enough to unmask minor traces indicating the hybrid status of novel species and, hence, clarify phylogenetic studies in which they are included. Second, the genome of *Arxula adenivorans* (dispersed 5S RNA genes group), in the process of analysis, exhibits a structure more closely related to filamentous fungi and may prove a very interesting basal species of the Hemiascomycetes.

The availability of new technologies has moved the critical steps in genomics from sequencing itself, that is no longer the limiting step, to upstream and downstream steps, namely the choice of relevant species or strains and the analysis of the flow of sequence data to unravel interesting biological questions. Yeasts, defined as fungal forms able to propagate indefinitely as unicellular organisms (mostly by budding), are observed both within the *Ascomycota* and the *Basidiomycota* the two main phyla of modern fungi or *Dikarya*. By extending our genomic exploration to more *Ascomycota* yeasts and by exploring *Basidiomycota* yeasts (poorly studied until now except for a few species), we anticipate approaching the question of unicellularity. In different *Basidiomycota* lineages, we observe the presence of closely related taxa of yeasts and filamentous fungi. This situation offers a nice opportunity to study by a comparative genomic approach some aspects of the evolutionary transition between unicellularity and multicellularity. In collaboration with laboratories from Belgium, Germany, The Netherlands and Spain, the Génolevures Consortium is now developing a new project called DIKARYOME with the Génoscope, which focuses on this aspect of eukaryotic genome evolution by sequencing a large set of new yeasts.

## 8. Conclusion

The association of French laboratories with a long-term interest in *S. cerevisiae*, was largely at the origin of comparative genomics of a eukaryotic branch by focusing on the study of a monophyletic lineage: the hemiascomycetous yeasts. The initial recommendation of the Génolevures Consortium for full-scale sequencing is now obvious with the appearance of the new-generation technology for DNA sequencing. The work of this consortium demonstrated that evolutionary routes are more diverse than expected from classical morphological criteria. It predicts that more genomic analyses conducted in the scope of comparative genomics will unravel plenty of unexpected biological novelties (hybrids, introgression, lateral gene transfer from bacteria, synteny conservation, integration of short mitochondrial DNA fragments into nuclear chromosomes, for examples).

The identification and the classification of living organisms, what we could name “biodiversity” is frequently discussed, even at the societal level but often in the sense of its reduction, but the tools required for its characterization are rarely discussed. The progress of molecular studies during the last fifteen years since *S. cerevisiae* was first sequenced, indicates that all genomes are subject to constant reshaping events, suggesting that only a minor

part of the potential biodiversity is detectable at a defined period. The recent progress of DNA sequencing technologies, at a rapidly decreasing cost correlated to huge increase of the data-flow, offers us a better possibility to observe nature.

As noted above, the introduction of DNA sequencing has changed the traditional identification and classification of yeast species based on morphologic, biological or biochemical tests. The DNA approach now paves the way to discover the extent and the roots of biodiversity with robust data. Focusing on the fungal kingdom only, a comparative genomic analysis based on the systematic sequencing of a significant number of species of the *Basidiomycota* phylum will offer numerous evolutionary landmarks to the scientific community to enrich the fungal tree of life or AFTOL (<http://www.aftol.org>). In addition, the new sequence data will be useful to characterize the extent of hybrids among previously defined species inside large collections. The Génolevures work has enhanced our understanding of eukaryotic genome evolution during the last ten years, it may contribute to better identify biological diversity in the following years.

## Disclosure of interest

The author declares that he has no conflicts of interest concerning this article.

## Acknowledgements

To all the members, past and present, of the Génolevures consortium, to Horst Feldmann (Ludwig-Maximilians-Universität, München), to André Goffeau (Université catholique, Louvain-la-Neuve), to the Génoscope, to the CNRS, to the French ‘Académie des sciences’, to the different institutions that provide the salaries of the scientists (INRA, Agro-Paris-Tech, CNRS, CEA, Institut Pasteur et les universités de Bordeaux 2, Lyon 1, Marseille 2, Paris 6, Strasbourg). This work was supported in part by funding from CNRS (GDR 2354, the Génolevures consortium) and in part by ANR (ANR-05-BLAN-0331, GENARISE).

## References

- [1] The Génolevures consortium, Genomic exploration of the hemiascomycetous yeasts, *FEBS Letters* 487 (2000) 1–149 (special issue, 21 publications).
- [2] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, et al., Life with 6000 genes, *Science* 274 (1996) 546–567.
- [3] O. Ozier-Kalogeropoulos, A. Malpertuy, J. Boyer, F. Tekaiia, B. Dujon, Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*, *Nucleic Acids Res* 26 (1998) 5511–5524.
- [4] C.P. Kurtzman, J.W. Fell, T. Boekhout, The yeasts a taxonomic study, (2011) Volume 1, part 1, Definition, classification and nomenclature of the yeasts, Elsevier, Amsterdam.
- [5] P.F. Cliften, L.W. Hillier, L. Fulton, T. Graves, T. Miner, W.R. Gish, et al., Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis, *Genome Res* 11 (2001) 1175–1186.
- [6] M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E.S. Lander, Sequencing and comparison of yeast species to identify genes and regulatory elements, *Nature* 423 (2003) 241–254.
- [7] F.S. Dietrich, S. Voegeli, S. Brachat, A. Lerch, K. Gates, S. Steiner, et al., The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome, *Science* 304 (2004) 304–307.

- [8] M. Kellis, B.W. Birren, E.S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*, *Nature* 428 (2004) 617–624.
- [9] B. Llorente, A. Malpertuy, C. Neuvéglise, J. de Montigny, M. Aigle, et al., Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*, *FEBS Letters* 487 (2000) 101–112.
- [10] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casarégola, et al., Genome evolution in yeasts, *Nature* 430 (2004) 35–44.
- [11] K.H. Wolfe, D.C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome, *Nature* 387 (1997) 708–713.
- [12] S.O. Suh, M. Blackwell, C.P. Kurtzman, M.A. Lachance, Phylogenetics of *Saccharomycetales*, the ascomycete yeasts, *Mycologia* 98 (2006) 1006–1017.
- [13] B. Dujon, Yeast evolutionary genomics, *Nature Reviews* 11 (2010) 512–524.
- [14] The Génolevures Consortium, *Genome Res* 19 (2009) 1696–1709.
- [15] P. Durrens, D.J. Sherman, A systematic nomenclature of chromosomal elements for hemiascomycete yeasts, *Yeast* 22 (2005) 337–342.
- [16] T. Martin, D.J. Sherman, P. Durrens, The Génolevures database, this issue.
- [17] T.W. Jeffries, I. Grigoriev, I.G. Grimwood, J.M. Laplaza, A. Aerts, A. Salamov, et al., Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*, *Nature Biotechnology* 25 (2007) 319–326.
- [18] Y.C. De Schutter, P. Lin, A. Tiels, S. Van Hecke, J. Glinka, P. Weber-Lehmann, et al., Genome sequence of the recombinant protein production host *Pichia pastoris*, a methylotrophic yeast, *Nature Biotechnology* 27 (2009) 561–576.