FEMS Yeast Research, 19, 2019, foz023

doi: 10.1093/femsyr/foz023
Advance Access Publication Date: 7 March 2019
Retrospective

RETROSPECTIVE

My route to the intimacy of genomes

Bernard Dujon*

Department Genomes and Genetics, Institut Pasteur, CNRS (UMR3525) and Sorbonne Université (UFR927), 25 rue du Docteur Roux, Paris F-78724 , France

*Corresponding author: Department Genomes and Genetics, Institut Pasteur, 25 rue du Docteur Roux, F-78724 Paris-CEDEX 15, France. Tel: +33-1-4568-8482; E-mail: bdujon@pasteur.fr

One sentence summary: Retrospective on the work of Bernard Dujon.

Editor: Terrance Cooper

ABSTRACT

Being invited by a prestigious journal to write the retrospective of one's life is first a great honor, and then a chore when starting to do it. These feelings did not spare me. But trying to recall my past to the best of my memory, I learned how lucky I was to have been born to a generation that witnessed so many scientific discoveries. There is little in common between the genetic courses I taught recently and those that I received more than 50 years ago. Thinking that a tiny bit of this fantastic evolution might come from my accidental encountering with yeasts is a stunning experience. I wish the same for the new generation.

Keywords: yeast; intron; homing-endonuclease; recombination; double-strand break DNA sequence; evolution; gene amplification; hybrid

I first entered a research laboratory in September 1969. It was in Gif-sur-Yvette, a green area south west of Paris where CNRS had established its first campus in 1946. I was 22, and just completed an advanced diploma in Genetics from University Paris 6, an avatar of the Faculty of Sciences of Paris that emerged after the 1968 student uprising. A mix of personal choices and lucky accidents brought me there. Half a century later, I acknowledge the privilege I had to have been able to make a living satisfying my intellectual curiosity. I hope that this rare freedom will persist for my grandsons, Hugues and Louis, now 7 and 2, as they make their own career choices.

FROM CHILDISH DREAMS TO A BAND OF SELF-MADE NATURALISTS

I became interested in Biology very young, a weird idea in a family concerned about my wasting time on things perceived as useless. Yet, as far as I can remember, I have always been more attracted by natural phenomena than by manufactured artifacts. I have always preferred history to fiction and the countryside to big cities. I remember my emotions at springtime when murmurs slowly grow in fields and forests, when naked woods bloom and landscapes swell in every shade of green until nature bursts in a multiplicity of forms and colors. As a child, I was thrilled.

It is perhaps why, from the 6th grade on, I started paying attention to lessons in the natural sciences. They were very basic, but delivered by excellent teachers at a junior college in Maisons-Laffitte, a western suburb of Paris where my parents settled in 1958. I was eleven and, with a few schoolmates, started collecting almost everything from our natural environment—plants, fossils, insects, shells, rocks—trying to figure out what they were, with the help of simple books. We even attempted to breed all kinds of small animals from nearby ponds and forests, not always welcomed by our parents. But we gradually learned a lot and, as years passed, an informal small club of self-made

Received: 4 March 2019; **Accepted:** 6 March 2019

Figure 1. 1963, Michel Arluison (left) and I at Maisons-Laffitte in front of his 'naturalist' cabin.

'*naturalists*' emerged, occupying a significant part of our free time after school (Figure 1).

ALONE WITH TEENAGE QUESTIONS

Maintaining an interest in natural sciences as a secondary school senior was not easy. Regarded as futile in the elitist French educational system, Biology was not even offered in most *curricula* for the *baccalauréat*. It was relegated to the less prestigious ones. There, however, with the help of excellent professors and a good library, I learned a lot and, in June 1965 while passing my baccalauréat, I also became a laureate of Biology at the *Concours général des lycées*, a nation-wide contest organized annually in France since 1747. Winning a difficult honorific competition obviously felt good, but what to do next?

At 18, my childhood was over and Biology offered limited professional careers. I had to make choices without much knowledge of their meaning. I knew little about the arcana of the French high education system, and nothing at all about research laboratories or institutions. Like many in their time, my parents had quit school at junior college level and nobody in our family or friends had any experience in higher education. I felt terribly alone, facing an unknown universe of uncertain issues, with many questions. I could try medicine, veterinary or agronomy, all offering respectable careers. But I felt more attracted by the secrets of natural phenomena than their practical applications. I decided to pursue a degree in Biology, without much idea of where it would lead, besides becoming a school teacher.

THE FACULTY OF SCIENCES OF PARIS—MORE STUDENTS THAN JOBS

In September 1965, I arrived at the *Faculté des Sciences de Paris*. Universities, as found in most countries, didn't exist in France until 1968. Ancient universities were abolished during the French Revolution and replaced by various specialized structures that evolved independently. In 1965, Faculties were places where, both teaching and research were carried out. I began with a smattering of mathematics, physics, chemistry, biology and geology. Lectures were excellent, inspiring and very motivating, despite the overcrowded amphitheaters and dilapidated premises. Professors warned us upfront that there were many more of us than available jobs in the future. Yet, I remained optimistic. It was intense work without much supervision. But this suited me well. I was used to working alone, I had done so for years, and my efforts were now paying off. I passed all exams scattered along this exhausting academic year and at its end, ranked in the top 1% of students. This is probably what determined the rest of my career because, with these marks, I was offered an opportunity to compete for an oral exam at *Ecole Normale Supérieure* (ENS).

ECOLE NORMALE SUPÉRIEURE—FROM ANONYMOUS BIOLOGY STUDENT TO NORMALIEN

I was not exactly sure what ENS was. Founded in 1795 as a place to educate professors for secondary schools, it became a prestigious multidisciplinary high school, training numerous university professors and French intellectuals. Very few students are recruited at ENS every year after severe written and oral competitions that most applicants prepare in special classes for 2–3 years after the *baccalauréat*. I ignored such preparations because they looked old-fashioned and boring to me. Now, only 1 year at the Faculty where science was much more interesting I was directly selected for the oral competition at ENS that I won, ranking first. At 19, I became a privileged *normalien* instead of the anonymous biology student that I was before.

The difference was enormous. The government supported my studies with a small salary for 4 years, and I had a much clearer understanding of where my curriculum would eventually lead me. I deeply measure my luck by terrible familial comparisons. At the same age, my father became a war prisoner in Germany, where he stayed 5 years. At the same age, my son François died. He was a young successful pharmacy student, entering his second year at the University after his *baccalauréat*. Why did he not have my chance? Where is the justice?

During my first 2 years at ENS, I attended lectures at the Faculty of Sciences in everything considered useful for the *aggregation* diploma. This is when I discovered Genetics. To me, genetics was different from other biological disciplines. It asked more fundamental questions about the very nature of life and provided clearer intellectual schemes to address them. I decided to specialize in it for my third year at ENS, a decision I have never regretted.

CNRS–A LUCKY START

For my fourth and last year at ENS, I had to make a difficult choice. As a *normalien*, I was expected to prepare the *aggregation*, the highest diploma for professors of secondary schools, which ensured a lifetime position in the French public education system. Instead, I chose to pursue a doctorate in genetics despite all of the uncertainties of research careers, even at that time. I entered the laboratory of Piotr Slonimski at the CNRS in Gif-sur-Yvette to work on yeast mitochondria. But a *Doctorat d'Etat* as existed at this time, was much more than a year of preparation. It required extensive work during which the applicant was supposed to obtain comprehensive original results and become an expert in the topic. How to support these additional years of research? I didn't really know. My young scientific life might have stopped there. But just as my ENS time ended, CNRS recruited me as a junior scientist, offering the best possible conditions to complete my thesis. What a privilege compared with the difficulties of today's students.

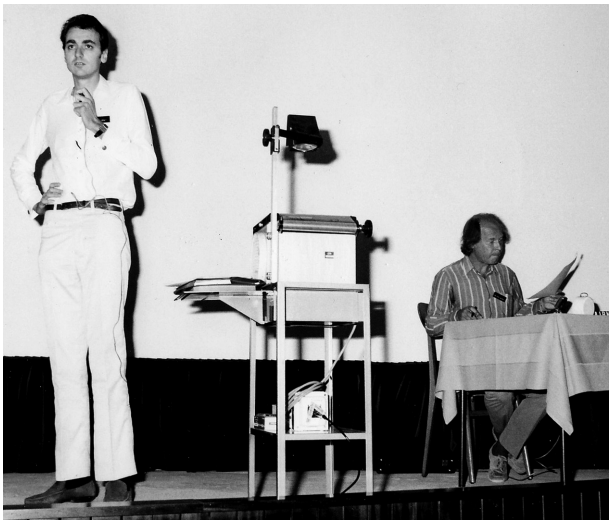

Figure 2. 1972, Presenting results of mitochondrial genetics at the sixth international conference on yeast genetics and molecular biology, Pisa, Italy. Piotr Slonimski was chairing the session.

THE BIRTH OF MITOCHONDRIAL GENETICS—MANY INEXPLICABLE RESULTS

The first mitochondrial point mutations, conferring resistance to erythromycin (E^R) or chloramphenicol (C^R), were isolated the year before I arrived in Slonimski's lab. Their ignorance of Mendelian rules was enough to pique my curiosity. Crosses of E^R or C^R mutants with wild-type yeast strains (E^S and C^S alleles) revealed entirely new rules of genetic segregation: alleles separated from each other during the first post-zygotic mitoses, forming distinct diploid lineages that showed no further segregation at meiosis. Furthermore, recombinants appeared, but in mysterious proportions. We had absolutely no idea about the genetic content of mitochondria. We knew they contained DNA, but we knew nothing about the genes carried—if any—or the form and complexity of the mitochondrial genome.

For my first 2 years of research, I performed crosses of yeasts studying the inheritance of these mitochondrial mutants in various conditions. P. Slonimski had built a team of six young scientists—including myself—rapidly joined by visitors from abroad. Numerous crosses were made, following the numerous hypotheses that proliferated from the accumulation of inexplicable results, themselves generating new inexplicable results, new hypotheses and new experiments.

Initially, mitochondria were supposed to have sex, explaining the mysterious recombination biases, an idea rapidly abandoned to the benefit of a third mysterious locus, designated *omega* with two alleles *omega*$^+$ and *omega*$^-$, affecting both E^R/E^S and C^R/C^S loci. Each yeast strain carried either one *omega* allele or the other. Recombination was observed in all cases but the frequencies of reciprocal recombinants became highly distorted when the two parental strains had opposite *omega* alleles. This became an important part of my doctoral work. I rapidly discovered that the *omega*$^-$allele was mutating to a neutral form, designated *omega*^{n} when selecting mutants at the C^R/C^S locus. But, strangely, no equivalent mutation was ever derived from the *omega*$^+$ allele (Figure 2).

A MODEL WITH A LITTLE HERESY

In 1974, with new mitochondrial mutations included in my crosses, I proposed a formal model of mitochondrial inheritance postulating that the interaction between *omega*$^+$ and *omega*$^-$ alleles (but not *omega*^{n}) initiated a sequential gene conversion reaching successively the C^R/C^S locus and then the E^R/E^S locus (but not the other loci), thereby replacing alleles originating from the *omega*$^-$ parent with copies from the *omega*$^+$ one (Dujon, Slonimski and Weill 1974). I had no idea of the molecular nature of the *omega* locus, but the quantitative predictions of this model fitted reasonably well the experimental data if one postulated, as I did, that the initiating event of the gene conversion was a double-stranded break at the omega$^-$ allele, a then heretical hypothesis for specialists of genetic recombination mechanisms!

I nevertheless continued my complicated calculations, trying to incorporate results from all of the crosses into my formal model and defended my thesis in 1976. But the nature of the *omega* locus remained mysterious. There was even a logical problem in terms of genetics. If *omega*$^-$ was the active allele and *omega*^{n} its inactive mutant, what was *omega*$^+$? If *omega*$^+$ was the active allele, what was the difference between *omega*$^-$ and *omega*^{n}? This convinced me that formal genetics was insufficient to reveal the nature of genes and genomes.

THE CRITICAL TRANSITION

Recombinant DNA and restriction enzymes had just been discovered. Molecular techniques were progressing quickly, opening entirely new fields. In 1977, modern methods appeared for DNA sequencing and, later the same year, introns were discovered. I was so excited that gene structures could now be examined and were revealing such surprises that I contacted Walter Gilbert at Harvard University about possible postdoctoral training in his laboratory. I told him about a hypothetical mobile-intron invading genes that I wanted to characterize by DNA sequencing. He appeared interested but said that it was only possible a year later because of space constraints. As a member of CNRS, it was not difficult to wait for this great opening and I spent the interval preparing for my visit.

The transition from my genetic results to the mobile intron that I told W. Gilbert about needs an explanation. Before completing my doctorate, I had tried to characterize the *omega* locus by deletion mapping using the *rho*$^-$ mutants. At this time,

François Michel, another student in Slonimski's lab, was trying to map mitochondrial DNA by partial thermal denaturation. Together, we had a sort of surrogate to the newly emerging gene cloning techniques. Our first results were, again, inexplicable until we recognized that, when compared to the *omega*$^-$ allele, the *omega*$^+$ allele had a large insert (*ca.* 1 kb) in the region of the mitochondrial DNA corresponding to the large ribosomal RNA gene. As soon as introns were discovered, the possibility that the *omega*$^+$ insert could be an intron in the large rRNA gene of mitochondria became an attractive hypothesis. But compared to the known introns of this time, it had two unique properties: it was optional despite lying within an essential gene and it was invading intron-less genes in crosses.

HARVARD UNIVERSITY—SEQUENCING THE FIRST MOBILE INTRON

I arrived at Harvard in September 1978 with my wife, Annie and our children, Cécile (4 years old) and François (1 $\frac{1}{2}$ years old). Settling a family in a foreign country is always challenging but the USA was an open-minded place. Originally planned for one year, we stayed in Cambridge, MA, almost 3 years. Gilbert's laboratory was an exceptionally creative place. Introns, gene structure, DNA cloning and sequencing were central. Many of the novel techniques to manipulate DNA and RNA emerged there, or were immediately adopted. I particularly benefitted from the last DNA sequencing methods of Alan Maxam.

I started sequencing the clones that I brought with me from Gif-sur-Yvette, a lucky break because in 1978 the city of Cambridge had banned the construction of recombinant DNA on its territory! I could only complete my work 6 months later using a P3 facility built for hazardous pathogens! Yet, by the fall of 1979, after a short summer break with my family in the eastern provinces of Canada, I had the molecular basis of the first known example of a mobile intron (Dujon 1980).

The importance of the exon sequences flanking the intron, as judged from the nature of my *omega*^{n} mutations, was identified in those bases. But the most surprising discovery was the presence of a long coding sequence within the intron itself. Translated using the mitochondrial genetic code that started to be elucidated, it could encode a lysine-rich protein of 235 residues. This was entirely unexpected. No intron was supposed to be coding. No such protein was detected before in yeast cells. If actually translated, what could be the function of such a protein? By analogy to transposable elements, a role in the capacity of this intron to invade intron-less genes was an attractive hypothesis. But how to demonstrate it with the available tools? I was thinking along this line when something else happened.

THE JUNGLE OF OTHER PUTATIVE INTRON-ENCODED PROTEINS

This was the time when I experienced a difficult period in my relationship with P. Slonimski. I had sent him all of my results by the end of 1979, before publishing them. He was working on introns of the mitochondrial *COB* and *OXI3* genes that, unlike the *omega* intron, showed no evidence of propagation in crosses. Instead, there were mutations in the introns that, based on complementation assays, he interpreted as affecting RNA-guides for splicing, an idea that he presented at Harvard a year before. Actually, in Slonimski's work, splicing was not directly determined at the RNA level but deduced from observed proteins.

Figure 3. 1980, the Gilbert's lab, at the announcement of the Nobel prize. In front of the Harvard Biolabs, Cambridge, MA. W. Gilbert on top of the bronze rhinoceros.

Interested in the molecular mechanisms of RNA splicing, Georges Church, a student of W. Gilbert, decided to reexamine this question directly as part of his PhD and, by the end of 1979, hypothesized that the splicing guides for the *COB* and *OXI3* introns (whose sequences were not available) were not RNA but intron-encoded proteins. They designated the putative intron-coded proteins as *spligases*. P. Slonimski immediately reacted by publishing the same idea, claiming his discovery of intron-encoded proteins and their function as *RNA maturases* with only a partial sequence of a *COB* intron and without any mention to the *omega* intron that was the only fully sequenced yeast mitochondrial intron at this time. There was obviously much more to be learned about introns (see below), but this episode about hypothetical intron-encoded proteins left me with a bitter taste.

A SECOND TRANSITION

I stayed at Harvard for another 18 months, uncertain about my return to France. I worked on replication-origins of yeast mitochondrial DNA in collaboration with Hugues Blanc, a former student at Gif-sur-Yvette who was now at Stanford. I wrote a major review on yeast mitochondrial genetics and I imagined future lines of research. In the fall of 1980, W. Gilbert received the Nobel Prize of Chemistry for his contributions to nucleic acid sequencing. This was an unforgettable experience for all members of his laboratory and I was truly fortunate to have lived such a moment (Figure 3).

It was during this period that I began to think about exploring the molecular diversity produced by natural evolution as an interesting complement to the experimentally based investigations. This was not a popular idea among molecular biologists who were focusing on the universal mechanisms of life regarded as fundamental because evolutionarily conserved. Diversity could only be of secondary interest, if any. Without a precise idea of what to do with them, I nonetheless started to collect various yeast species feeling that they might become important someday. The future demonstrated that I was right. Meanwhile, political changes in the USA convinced me to return to France.

BACK TO GIF-SUR-YVETTE –ADAPTATION, OLD LABS AND NEW DISCOVERIES

In May 1981, my family and I returned to Gif-sur-Yvette. The children, now 6 $\frac{1}{2}$ and 4, fluent in English, had to face the directive French schools instead of a self-confident American education. They apparently adapted. By September, they had completely dropped English and melted into their new environment but, retrospectively, I wonder whether they were not internally traumatized. Scientists are often too busy to pay enough attention to the small details that make lives.

I started my own team in an empty laboratory in the Centre de Génétique moléculaire du CNRS that P. Slonimski had reserved for my arrival. The facilities were inadequate for molecular genetics: narrow doors opening to sterile tiled rooms, a huge cold room, no bench and only one sink. They were made for people collecting microorganisms. I adapted, having experienced the Harvard Biolabs that had also never been renovated. The cold room was turned into a dark room where autoradiograms were manually developed, often ruining our trousers, while an old refrigerator was put back into service to replace the cold room. The sterile rooms became places for radioactivity, i.e. practically all our experiments, while homemade tables served as benches and desks. This is where we made great discoveries.

'BEAUX GÈNES SA'

Michel Hours, a laboratory technician, helped me to set up the room. F. Michel, finishing—his thesis, joined us immediately, and Alain Jacquier, a young, talented student started in September. H. Blanc, returning from Stanford, joined us a few months later. A young scientist from the Pasteur Institute, Pierre Dehoux worked with him on mitochondrial DNA replication. At six, the space was overcrowded, and more came in subsequent years, including Walt Fangman, on sabbatical leave from the University of Washington, Gertrud Burger and Franz Lang who, having resigned their Munich positions, were on the move to Canada, Elke Pratje and Carola Vahrenholz from Düsseldorf, Guillaume Cottarel, Laurence Colleaux and Mireille Betermier, three long-term students, and many other short-term visitors and students. The laboratory was constantly overcrowded to a point of ignoring elementary rules of safety. Gene-cloning companies were starting everywhere. In order to look fashionable we decided to put a sign on our lab door reading '*Beaux Gènes* SA' ('Beautiful Genes Inc.) After all, W. Gilbert had started *Biogen, Inc.* when I was in his laboratory, so why not *Beaux Gènes?* In his case, of course, it was not a joke (Figure 4).

BUSINESS AND COMPETITION

I started exploring the presence of the *omega* intron in the various yeast species I had collected at Harvard. To my surprise, I found it universally present in all of the *Kluyveromyces* species we tested but rarely in *Saccharomyces* species as if it were a *Kluyveromyces* intron in the process of invading *Saccharomyces* yeasts, a new idea at the time. Sequencing one of them with A. Jacquier, we found an internal coding sequence like that of *Saccharomyces cerevisiae,* an observation supporting the importance of their translation products. We also found that, in contrast, the remaining parts of the intron sequences aligned very poorly, suggesting distinct evolutionary forces.

This difference played an important role in our subsequent studies. F. Michel quickly found that, ignoring their coding sequences, the two *omega* introns of *S. cerevisiae* and *K. thermotolerans* could be folded into similar stem-loop structures, with compensatory sequence changes in stems and conserved sequences in their loops. Extending the same principles to five other yeast mitochondrial introns whose sequences were now available, he discovered that they could be folded into a common RNA secondary structure. This structure suggested how exon–intron junctions were defined during the splicing mechanism. We submitted a manuscript to *Biochimie*, but its publication was delayed for obscure reasons. Meanwhile, an English laboratory published very similar structures in a *Nature* article. We could never determine whether or not it was an independent work because our structures were presented publicly long before their publication. But it is satisfying to note that our *Biochimie* article has now received more citations than their *Nature* article! Perhaps because our article also contained a draft of a totally different secondary structure (not publicly revealed before publication) for two other yeast mitochondrial introns, indicating the existence of two distinct groups of introns. We designated them group I and group II introns, a nomenclature still in use today. Among group I introns, was the *Tetrahyma* intron that Thomas Cech had just discovered to be self-splicing. This suggested a direct implication of our predicted RNA structures in RNA catalysis. It is remarkable that these structures, proposed 35 years ago on theoretical grounds with very limited data, have now been proven correct by crystal structure analyses.

Meanwhile, Julius Subik had published a strain of *S. cerevisiae* that appeared neutral in mitochondrial crosses although derived from an *omega*$^{+}$ parent. This attracted my attention because, according to our model, the *omega*^{n} mutations that I had sequenced at Harvard would be incompatible with intron splicing. I contacted him and, being unable to work on this subject in Bratislava, he generously sent me the strain. Sequencing its *omega* intron, A. Jacquier and I discovered a frame-shift mutation in the coding sequence giving us the first direct demonstration that the intron-encoded protein was involved in the mobility of this intron. I called this mutation *omega*^{d}, for deficiency, and submitted the work to *Cell* in November 1984.

At this time, the laboratory of Ronald Butow in Texas was working on the same question using a complex technique to select mutants. *Cell* decided to publish our papers back to back. In our case, it meant an 8 month delay! A. Jacquier had to leave the laboratory for military duty and I had been nominated Professor at the University Pierre and Marie Curie the year before with a significant teaching load and numerous responsibilities. The competition seemed biased but, fortunately, back in 1983, I had decided to gamble.

A NEW TURN—MORE PRECIOUS THAN GOLD

Convinced that genetics was insufficient to elucidate the mechanism of the *omega* intron propagation, I wanted to purify the intron-encoded protein. Given its undetectable presence in yeast, its expression in a heterologous system appeared to me as the only way to proceed. However, numerous codons had to be altered to make the reading frame interpretable by the universal genetic code. In other words, I had to rewrite a gene of unknown function, with the hope of elucidating its function. No one had done anything crazy like this before. This was enough to convince me to do it! Fortunately, in a CNRS lab I could do experiments without first having to write grant proposals examined by expert panels. The difficulties were elsewhere. Oligonucleotide synthesis and *in vitro* mutagenesis were still in their infancy and

Figure 4. 1982, the early *Beaux Genes SA* team. Gif-sur-Yvette. From left, Hugues Blanc, Michel Hours, myself, Walt Fangman, François Michel, Alain Jacquier and Pierre Dehoux.

none of these techniques existed at the Centre de Génétique moléculaire.

By chance, I heard about Francis Galibert at Ecole Polytechnique that I had just joined as a part-time assistant professor in addition to my position at the University because, as a young professor, I could hardly support my family. My workload skyrocketed but this did not affect my motivation. I was simply working crazy hours. F. Galibert, who was working at Hôpital Saint Louis in Paris, had been trained in Fred Sanger's laboratory and liked ambitious, technically demanding projects. With his collaborator, Luc d'Auriol, they were the only ones in France able to synthesize oligonucleotides. When I explained my project to them, they were first surprised but decided to collaborate. Oligonucleotides were more valuable than gold! During the first half of 1984, they produced all of the synthetic oligonucleotides needed for my successive rounds of site-directed mutagenesis and by the end of this exhausting year, while Cell was delaying the publication of our work, I had changed 26 of the 235 codons of the *omega* intron reading frame and constructed an engineered gene to direct the synthesis in *E. coli* of a protein identical to the one expected to exist in yeast mitochondria.

RIGHT IN THE TARGET

A number of technical details had to be fixed in the subsequent weeks in order to monitor synthesis of a protein with such an unusual amino-acid composition. But I was rapidly rewarded for my efforts: a protein of the right size became visible on a gel and, using plasmids that I constructed for this purpose, a specific endonucleolytic cleavage of DNA was obtained in *E. coli*. The cleavage occurred specifically at the *omega*$^-$ locus, leaving the *omega*^{n} mutants intact, as expected. Clearly, the yeast *omega* intron-encoded protein was a DNA endonuclease generating a site-specific double strand break in the *omega*$^-$ locus. Exactly as predicted in my heretical model 12 years before! These results were obtained by May 1985, before *Cell* had eventually published our previous paper. It published my new results with less delay (Colleaux *et al.* 1986)

Figure 5. 1988, The Dujon family on the polar circle. Norway. From left François, Marc, myself, Cécile and Annie.

HAPPINESS AND SORROW

At this stage, I had every reason to be optimistic. A wonderful event had also happened at home. Our third child, Marc, was born in August 1984, starting a time of great happiness in a family of five. Perhaps the best birth present we received was from Michèle Kermorgan, a lab technician with whom I had collaborated. She offered me a sector cut in a circle of *Plexiglas* with a precise angle of 72 degrees, saying that I now needed a mold to cut cakes into equal pieces! She was absolutely right. Watching the interactions of three loved children as they grow and mature is immensely rewarding (Figure 5).

Our joy, however, was short-lived. In the summer of 1985, a terrible phone call informed me that H. Blanc had just died in a traffic accident on his way to the lab. From that moment on, the Beaux-gènes SA enthusiasm was irreversibly affected. Without H. Blanc, the laboratory became entirely focused on introns and their role in genomes. F. Michel concentrated on his modeling of group I and group II introns. For myself, during the time left by my teaching duties, I concentrated on my group I intron-encoded endonuclease and the mentoring of L. Colleaux, who had started her PhD.

THE MECHANISM OF INTRON HOMING

Despite my limited experience in biochemistry, I decided to purify the *E. coli*-produced protein in order to examine its enzymatic properties. This was again a bet. I had no indication that this protein could act alone *in vitro*. Furthermore, its very basic composition and strong tendency to aggregate did not facilitate my task. But, in the beginning of 1986, I finally obtained sufficient amounts of active fractions for *in vitro* assays on DNA and, with the help of Galibert's novel synthetic oligonucleotides, I could characterize the recognition site of the endonuclease. The result was astonishing! The site extended over 18 bp of DNA in the intron-less gene, spanning the intron insertion site. This immediately explained why the intron reading-frame could be expressed without cleaving its own DNA in omega$^+$ strains.

The results of my early crosses became clear. The cleaved intron-less gene was repaired using the un-cleaved intron-containing gene as template, hence making a novel copy of the intron that, in turn, produced more site-specific endonucleases, *etc.*... I called this process *intron-homing*. The beauty of this mechanism is its simplicity. It only requires the synthesis of a site-specific DNA endonuclease from a gene located within the recognition site of that nuclease to ensure the irreversible propagation of that gene in populations, using the general double-strand break repair mechanisms of the cells. This is one of the molecular mechanisms now recognized as *gene drive*. In total, it took me 15 years of research to arrive at this conclusion.

MY NEW DREAM

The limited internal degeneracy within the 18 bp recognition sequence of my intron-encoded endonuclease made it, by far, the most specific DNA endonuclease known at the time, potentially able to cleave DNA at a unique site within a large genome. It was the time when the human genome project was emerging and new procedures were needed to solve the problems of handling large DNA molecules. Next to the techniques of pulsed-field gel electrophoreses, jumping libraries and yeast artificial chromosomes, having a DNA endonuclease of unequalled cleavage specificity was for me the chance to enter an entirely new field of research.

But the Centre de Génétique moléculaire offered no chance to develop my lab in that direction. Further, I felt my presence there to be increasingly problematic given that my results did not coincide with Slonimski's *maturases*. In addition, new mobile introns were now appearing in other laboratories. Despite being less advanccd than *omega* in terms of molecular mechanism, their presence in other simple eukaryotes or bacteriophages made them potential competitors. I needed to act.

In January 1987, I received a call from Moshe Yaniv telling me that Institut Pasteur was looking for a yeast geneticist to set up a new research unit within its Molecular Biology department. The offer arrived just at the right moment for me. An advantage of the French public system of research, not easily understood abroad, is that positions are associated with a person, not with a working place. I could, therefore, move my lab to Institut Pasteur while keeping its affiliation to CNRS and my positions at the University Pierre and Marie Curie and Ecole Polytechnique. I decided to apply.

CREATION OF THE RESEARCH UNIT '*MOLECULAR GENETICS OF YEASTS*' AT INSTITUT PASTEUR

In November 1987, I moved my lab to Paris. My new environment was a loss in terms of surroundings: a compactly built urban area with traffic pollution instead of a gorgeous park by the banks of the quiet Mérantaise creek! But it was a gain in terms of scientific perspectives. Institut Pasteur wanted to reintroduce the yeast model into its activities, a topic essentially ignored there after the time of Louis Pasteur himself. To do so, I was offered a research unit of up to 12 investigators with complete freedom to organize its activities. A. Jacquier, who had just been recruited by CNRS, and L. Colleaux, who was completing her PhD, followed me. F. Michel, hating urban pollution, stayed in Gif-sur-Yvette. New collaborators arrived. Jeanne Boyer, Cécile Fairhead, Ettore Luzi, Claude Monteilhet, Arnaud Perrin, Anne Plessis, Agnès Thierry and our secretary, Martine Rambaud, formed the very first team of the young unit. Numerous others followed. I named my unit '*Molecular Genetics of Yeasts*', plural to indicate my still imprecise desire of not being limited to *S. cerevisiae* (Figure 6).

The unit, located in the Monod building, focused on introns and their protein products but also on DNA cleavage and repair in relationship to the novel ideas to explore genomes. The compactness of the building generated a highly collaborative atmosphere. Our exotic yeasts and bizarre projects piqued the curiosity of colleagues working on mouse development, cellular biology, bacterial genetics or enzymology, and we were rapidly adopted.

Intron splicing and RNA molecules occupied a significant part of our activities. A. Jacquier with Guillaume Chanfreau and other students focused on the *in vitro* splicing of group II introns. Pierre Legrain and Christine Chapon, returning from post-docs, and students concentrated on spliceosomal introns. Two years later, M. Rosbach paid a sabbatical visit to our unit with his wife Nadia Abovitch. Subsequently, Alan Tartakoff on sabbatical leave from Case Western Reserve University (Cleveland, OH) spent a year with us, studying nucleo-cytoplasmic transport of RNA molecules. I personally concentrated on group I intron-encoded endonucleases and their possible applications for genomics and genome engineering. I first wanted to improve the production of omega-endonuclease from *E. coli*. With the help of A. Thierry, a lab technician (now an engineer), we entirely synthesized an artificial gene using the codon preferences of *E. coli*. Another risky project. In 1988, this had never been done for a gene of that size. But we succeeded and it is this artificial gene that has subsequently been used for site-specific genome editing in various laboratories (see below).

HOMING ENDONUCLEASES—A NEW CATEGORY OF ENZYMES IS BORN

In 1989, as the characterization of additional mobile introns from distinct organisms progressed, it became clear that they

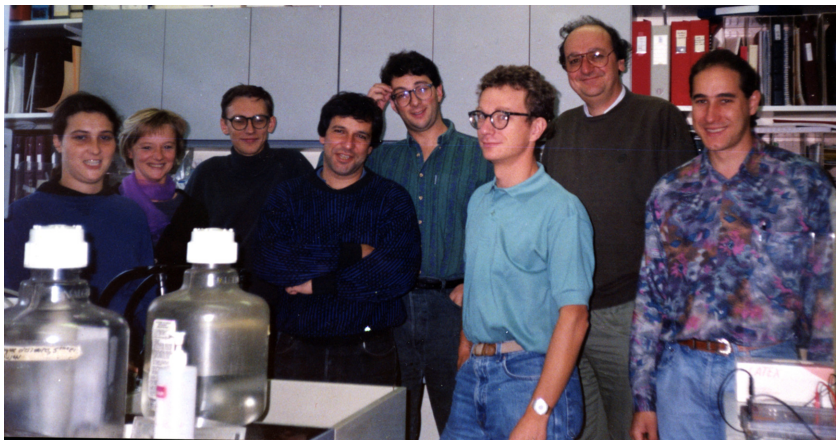

Figure 6. 1992, Some members of the young '*Molecular Genetics of Yeasts*' unit of Institut Pasteur. Paris. From Left: Cécile Fairhead, Agnès Thierry, Guy-Franck Richard, Alain Jacquier, Arnaud Perrin, Pierre Legrain, myself and Laurent Gaillon.

too encoded site-specific endonucleases (Dujon 1989). Following my *intron-homing* mechanism the nucleases became known collectively as *homing endonucleases* and designated individually following bacterial restriction enzymes with the prefix I, for intron.[1] My omega nuclease became I-*Sce* I. A new category of enzymes was born. But before homing endonucleases were named, C. Monteilhet a CNRS enzymologist who spent 2 years with us, L. Colleaux and A. Perrin, two PhD students that I supervised and myself had completely characterized the *in vitro* DNA cleavage reaction of I-*Sce* I produced by *E. coli*. Its asymmetric binding to the DNA recognition site followed by two successive single stranded cuts resulting in a double-stranded staggered cut with 4 bases 3' extensions and its low turnover number are important specificities to consider for applications of homing-endonucleases of this protein family (Figure 7).

AN AMBITIOUS INTERNATIONAL PROJECT

Shortly after arriving at Institut Pasteur, I received a call from André Goffeau from the University of Louvain-la-Neuve in Belgium. He had convinced the European Commission to undertake sequencing of the yeast genome. A huge task, the longest published DNA sequence then was less than 200 kb. The yeast genome was 50–100 times that size. A. Goffeau wanted to organize a network of European laboratories to start sequencing a small chromosome and thought about me as a former member of Gilbert's lab with experience in DNA. Considering that his request—sequencing ~10 kb in two year's time—was limited and could easily be handled by my lab in addition to its on-going projects, I joined the project.

WE CAN DO IT—CHROMOSOME III SEQUENCING

The European yeast genome program started officially in January 1989 with the goal of sequencing chromosome III coordinated by Steve Oliver (Manchester, UK). Werner Mewes from MIPS (Munich, Germany) was in charge of collecting the

Figure 7. 1988, Walking to a session at the Fifth Albany Conference on RNA, Rensselaerville, NY. From left, Rudolf Schweyen, myself and François Michel.

sequences of each participant—approximately 30 laboratories—and assembling them into a final chromosome III sequence. This was all just fine, except that no DNA had been prepared and physical mapping of yeast chromosomes—badly needed to organize the lab consortium—was still in its infancy. We worked with clones received from Maynard Olson (Washington University, Saint Louis, MO) and Carol Newlon (New Jersey Medical School, Newark, NJ). M. Olson had constructed a library of short yeast genome fragments cloned into bacteriophage lambda and deduced a first physical map of the shortest chromosomes. C. Newlon had cloned the fragments of a circular version of chromosome III into an *E. coli* plasmid. My lab was lucky to receive a clone from this second collection, much easier to manipulate than lambda clones in which the vector was four times larger than the insert. The insert we received was ~10 kb long, and C. Fairhead and A. Thierry sequenced it entirely using a random strategy. Although everything was done manually, the sequence was finished in less than a year in the middle of other activities.

[1] Numerous homing-endonucleases have now been described and several have been artificially designed. They belong to several families of protein structures that also include inteins (the prefix PI indicates protein-insert).

UNIQUE *IN VITRO* CLEAVAGE OF YEAST GENOMIC DNA

For me, sequencing a piece of chromosome III was originally a side-project. My main interest was to explore the possibilities offered by the exceptional DNA sequence recognition specificity of I-*Sce* I for whole-genome manipulations. This was not an easy task given that, beside *omega*$^-$ yeast mitochondrial DNA, no genome was known to contain an I-*Sce* I site and no simple tools existed to monitor a single cut within the DNA smear observed in gel electrophoreses of complex genomes. Yeast was the solution because artificial sites could be easily inserted at predetermined locations along its chromosomes and intact chromosomal DNA molecules were separable by pulsed-field gel electrophoresis. I, therefore, asked A. Thierry to construct transgenic yeast strains with I-*Sce* I sites inserted at intervals along their chromosomes and, with these strains, we could demonstrate for the first time the *in vitro* cleavage of an entire eukaryotic genome at a single site. Furthermore, the set of transgenic strains offered us the basis to develop and original physical mapping method of yeast chromosomes (*nested chromosomal fragmentation*) that played a key role in our participation to the yeast genome project (below).

UNIQUE *IN VIVO* CLEAVAGE OF YEAST GENOMIC DNA AND GENOME ENGINEERING

With the same set of transgenic strains, we could also demonstrate for the first time the possibility to engineer a eukaryotic genome in a site-specific manner. I had obviously no idea whether I-*Sce* I, expressed in the yeast cytoplasm from an artificial gene in a recombinant plasmid, would efficiently cleave chromosomal DNA within the nucleus and without secondary toxic effects to the cell,[2] a good enough reason to try. But the result proved immediately positive and A. Plessis and C. Fairhead engaged in numerous experiments to examine the effects of double-strand breaks along chromosomes in the presence or absence of homologous sequences in either allelic or ectopic locations on chromosomes, or even on plasmids. James Haber, on sabbatical leave from Brandeis University, and Miria Ricchetti, from another research unit, joined them in some of the experiments. These were the early days of site-directed genome engineering that, incidentally, also led us a few years later to the discovery that short fragments of mitochondrial DNA (NUMTs) were frequently inserted within healed nuclear chromosomes in the absence of homologous sequences, a potentially strong mutagenic force of evolutionary significance.

TUTZING MEETING

Members of the consortium organized by A. Goffeau first met by the misty banks of the romantic Starnberger See, south of Munich. We and three other labs had finished sequencing our piece of DNA. Others were struggling with various difficulties, mostly due to limited expertise with DNA. It was unclear how the program could be completed on time and S. Oliver had the difficult task of distributing the missing pieces in the absence of a precise physical map. Most participants were only interested in protein-coding sequences, not truly in the genome and, consequently, paid limited attention to completing the sequence. However, A. Goffeau, who had placed his scientific credibility on the project's outcome, was already looking for new coordinators to organize the sequencing of two additional chromosomes.

I volunteered for chromosome XI, and Horst Feldmann (University of Munich) volunteered for chromosome II. Our reasons were different. He had mapped chromosome II from a different strain and suspected that he could easily produce a mapped cosmid library from the lab strain S288c, chosen for the genome project. In my case, I wanted to apply my new I-*Sce* I mapping method to an actual large-scale project. The reason for choosing chromosome XI was simple: it was the most clearly separated chromosome from others on pulsed-field gel electrophoresis. It seems that no one understood this critical technical point. Some even started to speculate that I knew of an important secret gene on this chromosome!

When returning to my lab from Tutzing in November 1989, I was more than a year ahead in producing the mapped cosmid library needed to sequence chromosome XI. Most other contractors were still busy sequencing chromosome III. Chromosome III was published as a major Nature article in 1992 and, although behind schedule, made an important impact as the first eukaryotic chromosome ever sequenced.

KEEPING TRACK IS NOT EASY—JOB OF DNA COORDINATOR

It was very clear to me that cosmids were the only appropriate vectors to clone the yeast genome for sequencing because a precise physical map was needed. Although we had successfully sequenced random segments of our chromosome III fragment, shotguning a whole-genome was beyond imagination. In addition, it was incompatible with the organization of an international lab consortium. Instead, the idea was to construct a high-resolution physical map of the chromosome to minimize sequencing of overlapping segments. I made the necessary calculations to produce a high coverage cosmid library of the yeast strain to be sequenced.

As a geneticist, however, I couldn't imagine working with S288c itself. It lacked even a single genetic marker to distinguish it from possible culture contaminants and it was haploid, a status that *S. cerevisiae* hates. Crosses were excluded because we knew nothing about genetic polymorphisms between strains. I contacted Fred Winston from the Genetics Department at the Harvard Medical School (Boston, MA) who had constructed clonal derivatives of S288c with a few gene deletions to serve as genetic markers and with an inversion of mating-type. He generously gave me two haploid strains with complementary auxotrophies and opposite mating types that he called FY23 and FY73. I immediately crossed them to produce a homozygous diploid (except for the markers and the mating-type genes) that I called FY1679 (again, no secret, $23 \times 73 = 1679$). From this strain, A. Thierry constructed the cosmid library that eventually served for sequencing chromosomes VII, X, XI, XV and portions of others.

She then selected recombinant cosmids corresponding to chromosome XI (with a sufficient multiplicity to ensure complete coverage) and mapped them to I-*Sce* I intervals using our nested chromosomal fragmentation strategy and radioactive fingerprinting of *Eco*RI digests. The result was a complete physical map of chromosome XI at 3.7 kb resolution. Only the telomeres were missing. Ed Louis (Oxford, UK) cloned them independently using a new plasmid integration method. Sequencing of chromosome XI started, on schedule in January 1991. Over a hundred authors contributed to the work with the results

[2] The same questions remain today with the novel genome editing methods.

published as a major Nature article in 1994. It was the second eukaryotic chromosome ever sequenced, twice the size of the first (Dujon *et al*. 1994).

The next phase of the sequencing program had already started. I was in charge of chromosome XV, and Hervé Tettelin, a student of A. Goffeau (now Professor at the University of Maryland) stayed a year in my lab to coordinate the sequencing of chromosome VII. He and Meng Er Huang, working in the laboratory of F. Galibert, who coordinated sequencing of chromosome X, used our nested chromosomal fragmentation strategy for mapping the cosmid clones as we were using the same method on chromosome XV.

The work of a DNA coordinator was intensive. After distributing the verified cosmid clones to the participating laboratories, the coordinator followed advances, answered questions, resolved unexpected problems, made appropriate sequence assemblies in agreement with the physical maps and presented regular progress reports. A correspondent at MIPS helped run the sequence analysis programs and collected approved sequences.

SITE-DIRECTED GENOME ENGINEERING IN MAMMALS AND PLANTS

Despite its visibility and growing importance, the yeast genome project was only a part of my lab's activities. I wanted to extrapolate our *in vivo* results in yeast using I-*Sce* I (above) to more complex eukaryotes. We had no experience with animals or plants but experts in mouse cells were easy to find at Institut Pasteur. Jean-François Nicolas was one floor below us. There, an entrepreneurial young student, André Choulika, started his doctorate after a short rotation to my lab. He decided to apply I-*Sce* I to mouse cells with the hope to stimulate gene targeting by homologous recombination as we had observed in yeast. He built the molecular constructs to insert I-*Sce* I sites in mouse chromosomes while A. Perrin in my lab constructed the expression vectors with our artificial gene. The results were rapidly spectacular. Expression of I-*Sce* I in mouse cells had no noticeable toxicity and was directing insertion of *LacZ* at the selected site with a high efficiency. In today's terminology, it was the first example of genome editing in a mammalian genome. The work, patented by Institut Pasteur, was published in 1994 and 1995 and became one of the foundations on which A. Choulika later built the Cellectis company.

In 1996 I also collaborated with Holger Puchta (Barbara Hohn's lab in Basel, Switzerland) to express I-*Sce* I in engineered tobacco seedlings. Again, no toxicity was noted and significant stimulation of homologous recombination was observed at the DNA break. The I-*Sce* I system for gene targeting has subsequently been used in a wide variety of experimental systems: insects, nematodes, protozoans, algae, fungi. We received requests for our original artificial gene construct until recently.

THE FERNBACH BUILDING AT INSTITUT PASTEUR

In January 1995, Institut Pasteur decided to move the '*Molecular Genetics of Yeasts*' unit to a larger and more modern facility and to simultaneoulsy create an independent research unit on the same floor for A. Jacquier and P. Legrain to work on RNA. Two new senior members arrived in my lab, Odile Ozier-Kalogeropoulos and Fredj Tekaia, as well as a post-doc, Emmanuelle Fabre returning from EMBL and three students, Bertrand Llorente, Alain Malpertuy and Tereisa Teixeira-Fernandes. Almost all were working on genomics with the exception of E. Fabre and T. Teixeira-Fernandes who worked on nucleoporines and the architecture of the yeast nucleus (Figure 8).

Genome sequencing was not very popular at Institut Pasteur. Many were repelled by the repetitive nature of the techniques (ignoring that they were evolving rapidly) and, more fundamentally, by the idea that it was not hypothesis-driven research. They had simply forgotten the heuristic power of curiosity-driven research. Only three other laboratories, in addition to ours, were engaged in similar activities on *Bacillus subtilis* and on *Mycobacteria*. Furthermore, the results of the yeast genome were troublesome: many genes were of unknown and totally unpredictable functions, and redundancy was overwhelming. It was as if the very bases of molecular biology were suddenly jeopardized! People did not imagine that, if genomics is not hypothesis-driven research at the first genome, it becomes a very rich one at the second.

ITE MISSA EST

In 1995, the yeast genome program was in its last year. I was still coordinating the sequencing of chromosome XV but everything was in place for its completion. Except for unlikely accidents, it was foreseen that the yeast genome project was going to be a major success: the first eukaryote whose genetic makeup would ever be entirely deciphered (Goffeau *et al*. 1996). The completion of the yeast genome sequence was reported at a press conference in Brussels in April 1996, and the final meeting of the program took place in Trieste, Italy, in September of the same year (Figure 9). *Ite, Missa Est*[3] concluded A. Goffeau.

WHAT NEXT?

Before that, European participants to the sequencing consortium were thinking about functional genomics as a logical follow-up of their sequencing activities. Actually, under this vocabulary was hidden the desire of many to return to the biological problems they were interested in rather than to ask genomic questions. As a major contributor to the yeast genome program, I participated in many such discussions in 1995 and 1996. They were often complicated because, in reality, most people were not interested in the genome, they only regarded it as a large collection of genes among which were those corresponding to their topic of interest. The foreseeable result was obvious to me. We were trying to build a project in which a limited number of laboratories—the participants in the program—were going to do all the Biology in a short-term contract!

I regularly mentioned the lack of true genomic tools and the need to develop them, but our imagination on this point was limited. The consensus turned to the idea of building a collection of deletants covering, in principle, all predicted protein-coding genes and to perform a basic set of phenotypic assays on them. This could have been a great idea but, with the methods proposed, it was an extremely time-consuming project and, after calculation of its feasibility, it was decided to limit our ambition to sets of six genes per laboratory. Personally, I was not very attracted by the idea. Beyond the extremely complicated process put in place, a thousand laboratories were needed to cover the yeast genome!

[3] The translation is subject to numerous discussions. The concluding Latin words addressed to the people in the Mass of the Roman Rite, as will as the Lutheran Divine Service.

Figure 8. 1995, A sample of the growing culture of yeast geneticists at Institut Pasteur, Paris. From left, first row, Martine Rambaud, Jeanne Boyer, Micheline Fromont, Teresa Teixeira-Fernandes; second row, Cécile Fairhead, Agnès Thierry, Elise Dème, Monica Gnörich-Eck, Frank Lescure, Odile Ozier-Kalogerpoulos, Bertrand Llorente, Pierre Legrain, Fredj Tekaia; last row, myself, Laurent Gaillon, Jean-Christophe Rain, Arnaud Perrin, Guy-Franck Richard, Alain Malpertuy, Vladimir Manus and Emmanuelle Fabre.

Figure 9. 1996, The final meeting of the yeast genome sequencing project. Trieste, Italy. From left, front row: Ed Louis, André Goffeau, Werner Mewes, Howard Bussey, Karl Kleine; second row: Hervé Tettelin, Bart Barrell, Horst Feldmann, last row: Agnès Thierry, myself, Ron Davies, Mark Johnston, Steve Oliver and Meng Er Huang.

The project, however, was accepted by the European commission under the acronym EUROFAN and planned to start in January 1997 under the coordination of S. Oliver. I was in charge of several topics. But I was looking for something else. More than mere collections of genes, I suspected that genomes were hiding integrated dimensions that we needed to discover. The future told us that such dimensions were to be found in history more than in functions (see below). For the time being, I concentrated on paralogous gene families (with B. Llorente), large-scale deletions (with C. Fairhead), mini- and microsatellites (with G-F. Richard), overexpression (with J. Boyer and O. Ozier-Kalogerpoulos) and on building a comprehensive transcription map of chromosome XI in different life conditions (with C. Fairhead and G-F. Richard). One of the very first of its kind.

SILENT SPRING

While this was going on, my son François was experiencing internal difficulties that I didn't understand. Despite my close relationship with him, at least as I thought, I was unable to evaluate the growing intensity of his despair as time passed. I felt

Figure 10. 2011, a break at the ITMO conference, Paris. From left, myself, André Goffeau and Jean-Luc Souciet.

terribly useless. I obviously did not do what I should have. And in the beginning of January 1997, in the darkest period of the year, after an apparently normal day, my son poisoned himself in his room at home. Peacefully, as he always behaved. The hospital emergency proved unable to save him. My family was devastated. My daughter Cécile was engaged in her medical studies and needed help. My younger son Marc was at college and looked courageous. My wife, Annie, was striving to survive for all of us in this blackness. She told me to continue.

She was right and I did. Two weeks later, I gave talks at various places as previously planned. I don't know if the audience noticed but months later I was still unable to concentrate. I even went to Hilton Head, SC, for an international meeting on small genomes, at the very same place where I had discussed about my children with A. Goffeau few months earlier. I went to Stellenbosch, South Africa, for the XVIIIth International Conference on Yeast Genetics and Molecular Biology. And several other places, as usual. Annie started to work part-time as a librarian. She always loved books. For myself, I had a heavy teaching load at the University, and major responsibilities as president of the Institut Pasteur scientific council. I don't remember what I did in the lab. This was also the time when the French sequencing center, Génoscope, opened under the direction of Jean Weissenbach and I was nominated to join its scientific council.

Time passed. Genomics was accelerating with novel genomes and imaginative methods for functional analysis. C. Fairhead replaced me on several occasions in the EUROFAN project. A small relief came with my nomination to the *Institut Universitaire de France*, a selective delegation established to help a small number of university professors focus on their research by reducing their teaching duties. We spent a summer month in Ireland, where authenticity helped healing our wounds. The fall arrived. Followed by the darkest days of the year. Again.

Two new students, Gwenaëlle Badis-Bréard and Gaëlle Blandin, arrived to work on the functional genomics of *S. cerevisiae*. Reality was striking back. In February 1998, Jean-Luc Souciet, a colleague of the yeast genome project, invited me to spend a week in Strasbourg, hosted by the University, where I started to think projects anew.

THE DAWN OF YEAST COMPARATIVE GENOMICS

In 1998, DNA sequencing, despite automation, remained a costly operation, limiting its applications. Having one reference genome for a major group of organisms was considered sufficient. The yeast genome was now finished and, most people focused on functions—*post-genomics* as they liked to call it—without concern for the genomes themselves. The work that J. Weissenbach started at Génoscope illustrated the opposite. He understood the heuristic power of comparative genomics and was sequencing the genome of *Tetraodon* with the goal of facilitating interpretation of the human genome.

As we met in Brussels for administrative reasons, I asked him what he would think about exploring the genomes of other yeast species. As usual, his immediate reply was direct: what is a yeast species? Honestly, nobody knew and it is even less clear today. But this was not my point. I raised the issue because with less than 600 Sanger reactions on random pieces of DNA from *Kluyveromyces lactis*, a yeast differing enough from *S. cerevisiae* to look interesting, A. Malpertuy and O. Ozier-Kalogeropoulos identified nearly 300 protein-coding genes by sequence comparison with the predicted proteins of *S. cerevisiae*. This seemed to me like an efficient way to estimate the actual divergence between yeasts.

At year's end, J. Weissenbach informed me that Génoscope could offer 50 000 Sanger reads to explore yeast genomes. By today's techniques, such numbers are meaningless; reads are now counted in billions. But this was a hundred times what had been done in my lab. I needed only write a small project proposal. I called J-L. Souciet and my old friend from ENS, Claude Gaillardin, who both had expertise with non-conventional yeasts, and the three of us contacted colleagues we knew working on other yeasts than *S. cerevisiae*. Michel Aigle in Bordeaux, Monique Bolotin-Fukuhara in Orsay and Micheline Wesolowski-Nouvel in Lyon accepted to join the crew and, after discussions, we selected the most appropriate 'yeast species'.

RANDOM SEQUENCE TAGS

Thirteen yeasts were chosen, suspected to represent a broad sample of evolution within the Hemiascomycetes, a subphylum of Ascomycetes now called *Saccharomycotina*. The six labs shared preparation of the genomic libraries. We chose a random 3–5 kb DNA fragment strategy whose sequencing from both ends would maximize information about neighboring gene conservation (microsynteny). Using LiCor sequencers, Génoscope produced very long (*ca.* 900 nucleotides) high quality Sanger reads. This was perfect for our project. Each read was translated in six frames and compared with predicted *S. cerevisiae* proteins using a visualization routine written by F. Tekaia.

WHAT A HARVEST–20,000 NEW YEAST GENES

Over 40 million nucleotides of DNA sequence! We had never had so much yeast genomic data before, and didn't know what we'd discover. The project looked appropriate for the training of students. Each of them was in charge of one or two yeast species under the supervision of a senior scientist. In my lab, G. Blandin, annotated *Pichia angusta* (now *Ogataea polymorpha*) and *Candida tropicalis*, B. Llorente, *Kluyveromyces marxianus* and A. Malpertuy *Kluyveromyces thermotolerans* (now *Lachancea thermotolerans*). Pascal Durrens a CNRS scientist from the Aigle's lab joined the team of David Sherman at the LABRI (the CNRS center for informatics and applied mathematics in Bordeaux) to organize a place to store the data and the analyses.

Figure 11. 2011, The 10th anniversary of the Génolevures Consortium. Under the Coupole of Institut de France, Paris.

Figure 12. 2016. Doctor Honoris Causa from University of Perugia, Italy. To my right on the picture, Prof. Gianluigi Cardinali, promotor of the ceremony, and (with the ermine mantle) Prof. Francus Moriconi, Magnificent Rector of the University.

Everything was falling into place. My membership in the Institut Universitaire de France reduced my teaching load and, as the University was reorganizing its curricula, I was able to create a novel course for last year undergraduate students, specializing on genome analysis. This course has developed successfully in the continuing years, and has populated numerous labs with students interested in genomics. Also, the EUROFAN project was entering its second phase. I was in charge of more integrated approaches to functional genomics, closer to my interest. But comparative genomics occupied a growing part of my agenda.

Our first exploration of yeast genomes was amazing. About 20 000 new yeast genes discovered, three times the number that previously existed in all databases and we learned about sequence divergence from *S. cerevisiae*, conservation of microsynteny, gene redundancy and functional categories. Retroactively, we also had a much clearer understanding of the *S. cerevisiae* genome itself. A new field was opening that no one else had explored before. Some yeasts were related to *S. cerevisiae*, others were very different and probably very distantly related. Thus, the *Saccharomycotina* represented a much larger evolutionary span than anticipated. The results were published in a special issue of FEBS letters edited by H. Feldmann and illustrated by his drawing of Montmartre when he was a young student in Paris. The issue contained 21 articles under the title *'Génolevures: Genomics exploration of the hemiascomycetous yeasts'* It appeared December 22nd, 2000, 9 days before the end of the XXth century. EUROFAN was over. Its final meeting took place in the University of Salamanca, Spain, in a building dating from the XVth century. We no longer had a common project but we had many friends across the European Union.

Figure 13. 2016. Celebration day at Institut Pasteur, after my retirement, Paris.

THE GÉNOLEVURES CONSORTIUM–WE NEEDED MORE

CNRS, hearing the results of our collaboration with Génoscope (the six labs were affiliated with CNRS), decided to offer us a renewable 4-year contract under the name *Génolevures* that J-L. Souciet agreed to coordinate. The contract actually lasted 12 years and was subsequently continued by a new *iGénolevures* contract coordinated by C. Fairhead, now professor at University Paris-Saclay in Orsay. Gilles Fischer, a post-doc from Ed Louis' laboratory (Oxford University) interested in our new genome data joined my lab. A new student, Romain Koszul, arrived shortly later and started to work on mitochondrial DNA of the new yeasts, reminding me of my younger years. Institut Pasteur remodeled its research departments on thematic bases and I became head of the genomics department.

The Génolevures consortium started in January 2001. For people not familiar with the French research system, note that CNRS supported coordination of the consortium but not its operating costs including those of DNA sequencing. Génoscope was devoted to the human genome. We could only concentrate on existing data, trying to make new discoveries from their analysis. Two new post-docs with training in informatics joined the lab: Emmanuel Talla, who had worked on ATPase with A. Goffeau (Louvain-la-Neuve) and Ingrid Lafontaine who had worked on the analysis of DNA sequences with Richard Lavery (IBPC in Paris). We needed new data if their expertise was to be applied to the benefit of yeast genomics.

I was lucky to benefit from the small sequencing facility that Institut Pasteur had just established. I decided to start sequencing *Candida glabrata*, a important human pathogen that had attracted much less attention than *Candida albicans* to which it is not related (the word *Candida* only indicates a yeast whose sexuality is unknown). *Candida glabrata* is a member of the *Nakaseomyces* genus, related to *S. cerevisiae*. Christophe Hennequin, an MD from Amiens University Hospital, who spent a year in my lab to explore pathogenic yeasts, had brought my attention to this yeast. We decided on whole-genome shotgun sequencing but, for technical reasons, limited our ambition to 3X coverage (approximately 120 000 Sanger reads).

Meanwhile, the Sanger center sequenced *Schizosaccharomyces pombe*, the only other complete yeast genome that became available after *S. cerevisiae*. The two yeasts, however, have very little in common. *Schizosaccharomyces pombe* belongs to the *Archiascomycetes*, another subphylum of *Ascomycota* now called *Taphrinomycotina*. We learned later that the rare common features between the two yeast genomes were evolutionary convergences rather than conservations from shared ancestry. In 2002, after completion of the human genome, Génoscope informed us of its renewed ability to offer large-scale sequencing to the scientific community and asked me to chair its new scientific council. Capillary electrophoreses became available, increasing throughput and decreasing costs.

FOUR NEW YEAST GENOMES

The Génolevures consortium proposed sequencing the genomes of *Kluyveromyces lactis, Debaryomyces hansenii* and *Yarrowia lipolytica*, and completing the *C. glabrata* genome (*i.e.* 4 times more sequencing than we had done). The project was accepted, and resulted in a major Nature article (Dujon *et al.* 2004). This same year, the genomes of three other *Saccharomycotina* species were reported, each compared to *S. cerevisiae*, but Génolevures alone reported four new ones and made the first multidimensional comparisons. The consortium established the first catalogue of yeast protein families and examined their variation between species. It also engaged in analyses of the yeast sequences involved in replication, repair, recombination, mating and meiosis as well as sequence repeats, tandem gene arrays, tRNA genes

and the evolution of the genetic code, introns, pseudogenes, NUMTs, map rearrangements or specific chromosomal structures like telomeres and subtelomeric regions. Numerous publications followed.

SPECIFIC PROJECTS AND MORE YEAST GENOMES

Several specific projects continued in my lab in parallel. Interested in gene dosage compensation, R. Koszul and G. Fischer later joined by Cécilia Payen, imagined experiments in *S. cerevisiae* that resulted in the discovery of large segmental duplications whose mechanisms of formation became their major research line, Similarly, E. Fabre and Pierre Thérizols, discovered the importance of telomere tethering using double-strand break assays in *S. cerevisiae*, while G-F. Richard, Alix Kerrest and Valentine Mosbach concentrated on microsatellite stability. But yeast genomics had a growing influence on my lab. In line with my own interest on mitochondria, Christine Sacerdot who joined us in this period, used the NUMTs to monitor the recent evolutionary events in several yeast clades. *Candida glabrata* became as important as *S. cerevisiae*. C. Fairhead, with Héloise Müller, started to develop it as a genetically tractable yeast as early as 2003. She then coordinated the first comparative genomic analysis of the *Nakaseomyces* genus that includes several pathogens. The *C. glabrata* genome was also instrumental in our discovery of long coding repeats, that we called *megasatellites*, in genes involved in cell adhesions and pathogenicity.

As it became clear that the evolutionary range of *Saccharomycotina* was much broader than anticipated, additional sequences were needed in new lineages. We asked Génoscope to sequence *Kluyveromyces thermotolerans* (now *Lachancea thermotolerans*), *Zygosaccharomyces rouxii* and *Pichia sorbitophila* (now *Millerozyma sorbitophila*), plus a novel species of biotechnological interest, *Arxula adeninivorans* (now *Blastobotrys adeninivorans*). New participating laboratories joined the consortium and, Génolevures became an important training platform for students and young scientists, many of whom have now established successful laboratories that continue to work on yeasts.

THE ORIGIN OF GENES

Following courses that I gave in 2003 I wrote a small book about gene evolution. My interest shifted from the observed molecular diversity of yeast genomes to the mechanisms of genome evolution. In particular, I was fascinated by the question of how do genes arise beyond the then classical view of duplication-divergence that followed S. Ohno's important ideas. I remembered the exon-shuffling hypothesis that W. Gilbert had proposed several years earlier and could not stop thinking about the role of RNA molecules. In addition to looking for clues in the near dozen yeast genomes that became available, I wanted to engage in evolutionary experiments. I was particularly fascinated by the fact that paralogous families of genes for non-coding RNA molecules—especially tRNA genes—seemed to behave differently during genome evolution from the paralogous families of protein-coding genes. This is still an open question.

THE DIRECTION OF INSTITUT PASTEUR—WATCHING OVER THOUSANDS

In 2005, the new General Director of Institut Pasteur, Alice Dautry, asked me to join her as Associate General Director in charge of the scientific direction. This was an interesting challenge, and a proof of confidence. More than 2000 people worked at Institut Pasteur, in over 140 laboratories or services. I found myself again with a completely crazy agenda, having to supervise research departments, technical platforms, teaching activities, medical services, *etc.* and to make proposals in the rapidly changing environment of this period. I stopped teaching at the University, canceled several other responsibilities and severely reduced my activity in the lab.

In charge of the relationships of Institut Pasteur with the Universities, I also became a member of the scientific Directory of University Pierre and Marie Curie where more than 12,000 people and 30,000 students were working. I created a new laboratory in the University in which mathematicians could collaborate with biologists on various questions of quantitative biology, logical modeling and genomics. G. Fischer from my lab, followed by I. Lafontaine, established his research group there. At Institut Pasteur, I succeeded in reorganizing the research departments on more logical scientific grounds, recruit new PIs, develop technical facilities and start a new international PhD program. But, after nearly three years, during which I learned a lot, I continued to be more attracted by science than by its management.

PROTOPLOID *SACCHAROMYCETACEAE*

When I returned to my lab, the yeast genome sequences of Génoscope had not been completely exploited and, with the emergence of the new solid-state sequencing technologies, I anticipated accelerating competition. Two years later, the genome sequences of more than 20 yeast species were available (Dujon 2010). We focused on *L. thermotolerans*, *Z. rouxii* and *Saccharomyces kluyveri* (now *Lachancea kluyveri*) whose sequence had been offered to us by Mark Johnston (Washington University, Saint Louis, MO) for annotation and comparison. With these data and the previously published *K. lactis* and *Eremothecium gossypii* genomes, we had a rich description of five species of the *Saccharomycetaceae* family that, unlike *S. cerevisiae* and *C. glabrata*, were not descendants from the postulated whole-genome duplication. They were, therefore, designated '*protoploid*'. These protoploid species illustrated the basic genome architecture and gene content of *Saccharomycetaceae* yeasts (Figure 10).

A new student, Thomas Rolland with an informatics background, arrived in my lab at this time. He started to work on the conservation of syntenic blocks in our five protoploid yeasts. Numerous rearrangements were observed between the species, leaving blocks of syntenic orthologs whose size distribution argued against random chromosome breakage, in agreement with our previous suspicion of lineage-specific rates of rearrangements. His closer examination of the syntenic blocks also revealed the presence of horizontally acquired genes from bacteria, a phenomenon now recognized as frequent in nearly all yeast genomes, contributing to their physiological diversity.

INTERSPECIES HYBRIDS AND RETICULATED EVOLUTION

The next yeast, *M. sorbitophila,* revealed a new surprise. Its genome harbors seven pairs of chromosomes, some being homozygous (less than 0.01% sequence divergence between the homologs), the others heterozygous (up to 15% sequence divergence between the homologs). Such an unusual genome had never been observed before. They are, however, frequent and

result from the fact that diploid yeasts or hybrids tend to homogenize pieces of their chromosomes by a process called *loss of heterozygosity* (LOH) resulting from aborted meiosis or break-induced replication during clonal growth. After LOH, a large piece of chromosome or even the entire chromosome is converted into a copy of its homolog, duplicating the genetic information of one parent while eliminating the other. Consequently, clonal cultures of heterozygous diploid yeasts rapidly yield heterogeneous populations of chimeric homozygous genomes, maximizing the genetic diversity on which selection can act. This was not known in classical textbooks of genetics. It is, however, an essential phenomenon in evolution and, possibly, in development as well.

But *M. sorbitophila* also illustrated the fact that many isolates of yeasts are now known to represent the progenies of interspecies hybrids in various stages of evolution. Rather than genetically separated entities, the species appear as stratified populations having decreasing probabilities of genetic exchanges with other populations. Divergence-based incompatibilities, as postulated long ago by Bateson, Dobzhansky and Müller to explain speciation, remain exceptional. Instead, the phenotypic innovations brought by horizontal acquisitions are prone to favor the emergence of new lineages. The recent discovery by Lucia Morales, the last student that I mentored, of a cluster of nitrate assimilation genes of plant or fungal origin in the genome of *Kuraishia capsulata* and related species exemplifies this point. Similarly, the recent discovery by Toni Gabaldon's lab that the whole-genome duplication recognized by Ken Wolfe (Dublin, Ireland) in the ancestry of *S. cerevisiae,* followed an earlier hybridization between two diverged clades of protoploid *Saccharomycetaceae* indicates the importance of ignoring species boundaries during evolution (Figure 11).

STRUCTURAL MUTATIONS

In 2005, when I was thinking about the evolution of paralogous genes for non-coding RNAs, I could not predict where it could lead me several years later. At the time, I asked A. Thierry to replace essential *S. cerevisiae* genes coding for amino-acyl-tRNA synthetases (RS) with their orthologs from *Y. lipolytica*, the most distantly related yeast genome we had, with the idea of simultaneously inserting cognate tRNA genes from *Y. lipolytica* to monitor their replacement of the genuine *S. cerevisiae* tRNA genes. To my surprise, however, some RS replacements were viable without the cognate tRNA molecules because the *Y. lipolytica* RS were able to charge the *S. cerevisiae* tRNA molecules with a sufficient efficiency for survival. The strains were severely unfit but, when cultivated in rich medium, nearly normal growing mutants appeared whose genome sequences revealed massive amplifications of long chromosomal segments bearing the transgenes. The increased production of foreign RS was sufficient to ensure the charge of *S. cerevisiae* tRNA molecules to levels supporting nearly normal growth rates. The amplifications were stable despite significantly increased DNA content in the yeast nucleus, indicating the flexibility of genome size. Further surprises came after Varun Khanna, a young bioinformatician that joined my lab, found that instead of the classical homologous recombination between repeated sequences, the amplifications resulted from rolling-circle types of replication initiating at replication origins and terminating in the partially repeated sequences of the genome, creating a variety of aberrant chromosomal structures compatible with life and bearing great evolutionary potential.

EPILOGUE: THE 15 BUDS OF MY LAB

The serendipitous episode that concluded my life in a research laboratory is, perhaps, the best reward that a scientist may dream of. According to the rules of French Universities, I became an emeritus professor on September 1st, 2015, after my 68th birthday. The '*Molecular Genetics of Yeasts*' Research Unit of Institut Pasteur terminated, all its members now working elsewhere. Over the 28 years of its operation, more than 120 people have worked in the unit. Altogether, we produced more than 250 scientific articles and 22 PhDs and Habilitations. More importantly, we had an enjoyable time. More than 800 colleagues from all parts of the world have published with me. Some of them, unfortunately, are no longer with us. It is obviously impossible to give credit to all of them individually, but I would like to express my most sincere acknowledgments and ensure them of my highest esteem.

Yeast genomics is continuing worldwide at an accelerated pace. It is amazing to see the publications that appear. I wish to be able to continue following them for a long time (Figure 12). As a French academician, I am supposed to be immortal, but the Academy is unfortunately still working on the protocol. Nearly all of those who passed in my lab have continued in biological research or related fields, many of them obtaining successful academic positions while in my lab or shortly thereafter Alphabetically: G. Chanfreau (UCLA); L. Colleaux (Institut Imagine, Hôpital Necker-Enfants maladies, Paris); A. Jacquier (Institut Pasteur); R. Koszul (Institut Pasteur); E. Fabre (Hôpital Saint Louis, Paris); C. Fairhead (Universiy Paris-Saclay in Orsay); G. Fischer (Sorbonne Université, Paris); I. Lafontaine (Sorbonne Université, Paris); B. Llorente (Aix-Marseille Université); F. Michel (Gif-sur-Yvette); A. Plessis (Université Paris Diderot, Paris); G-F. Richard (Institut Pasteur); E. Talla (Aix-Marseille Université); H. Tettelin (University of Maryland) and T. Teixeira-Fernandes (Sorbonne Université, Paris).

My lab has budded 15 times (Figure 13).

Conflicts of interest. None declared.

REFERENCES

Colleaux L, d'Auriol L, Betermier M *et al*. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* 1986;**44**:521–33.

Dujon B. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the ω and rib-1 loci. *Cell* 1980;**20**:185–97.

Dujon B. Group I introns as mobile genetic elements: facts and mechanistic speculations - a review. *Gene* 1989;**82**: 91–114.

Dujon B. Yeast evolutionary genomics. *Nat Rev Genet* 2010;**11**:512–24.

Dujon B, Alexandraki D, André B *et al*. Complete DNA sequence of yeast chromosome XI. *Nature* 1994;**369**:371–78.

Dujon B, Sherman D, Fischer G *et al*. Genome evolution in yeasts. *Nature* 2004;**430**:35–44.

Dujon B, Slonimski PP, Weill L. Mitochondrial Genetics IX: a model for recombination and segregation of mitochondrial genomes in *Sacch a romyces cerevisiae*. *Genetics* 1974;**78**:415–37.

Goffeau A, Barrell BG, Bussey H *et al*. Life with 6000 genes. *Science* 1996;**274**:522–67.