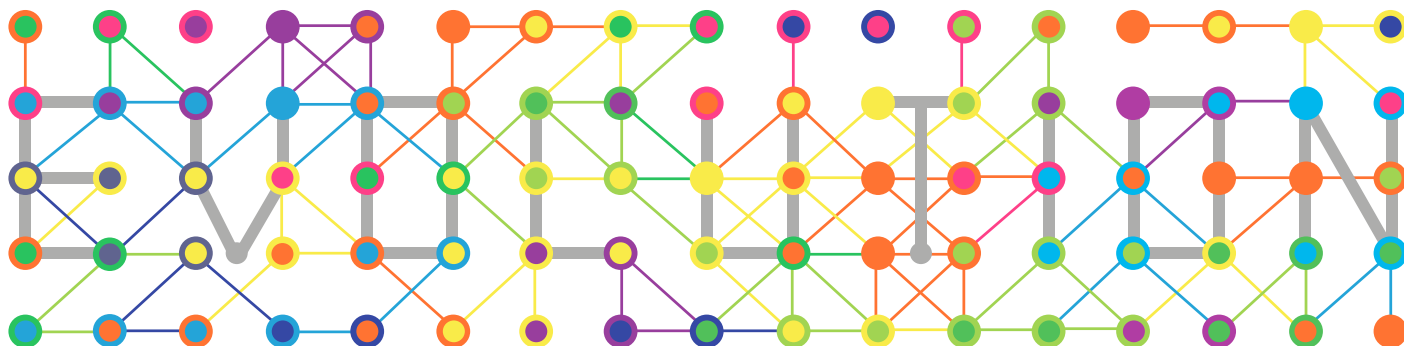




Understanding the past to predict the future

Expert in palaeogenomics **Dr Jérôme Salse** provides an insight into this fascinating discipline, which can reveal how species evolved by analysing ancestral genome sequences. He also discusses its translational potential to identify the genes behind biological processes key to agriculture and medicine

For those who are unfamiliar, could you briefly explain the field of palaeogenomics?

Palaeogenomics is the study of the evolution of living species through the analysis of ancestral genome sequences. There are two methods of analysing ancestral genome sequences, the first of which comprises mathematical reconstructions of ancestral genomes from comparisons of modern genomes. The second approach involves actually sequencing ancient DNA. Both techniques are performed in our palaeogenomics laboratory at the National Institute for Agronomic Research (INRA) in Clermont-Ferrand, France.

The major objective of palaeogenomics is to decipher how modern species have evolved over time. Comparing the genomes of great apes and modern humans, for example, allows us to identify the major gene functions associated with our most recent evolution. Another key aspect of palaeogenomics concerns understanding the impact of human domestication and selection on crop plant and farm animal genomes. Comparing the ancestral genomes of domesticated species allows us to identify genes that have been selected by humans for key traits (such as enhanced productivity or resistance to disease), and then to propose additional diagnostic sequences for future selection programmes.

How can palaeogenomics help us to better understand evolution?

Reconstructed palaeogenomes enable us to determine the relationships between

modern species. We can investigate the evolutionary fate of ancestral genes or genomes by identifying the changes (for instance, chromosome fusion, fission and translocation, as well as gains or losses of genes) that have taken place between ancestral and modern genomes, and then assign those changes to specific lineages or species.

What are the principal aims of your research?

We can retrace the evolution of species by comparing their genome sequences – if two species are closely related, they will share similar sequences. Our objective is to unveil the evolutionary mechanisms that created such diversity in the genomes and ultimately in the biology of plants and animals.

Why is there a need to compare the genome sequences of these organisms?

There are two main objectives when comparing animal and/or plant genome sequences – one fundamental and one applied. The fundamental scientific objective in performing such studies is to understand the evolution of modern living species over the last 500 million years. The applied palaeogenomics objective is transferring knowledge from model species (eg. mice or apes for animals and *Arabidopsis* for plants) to more complex species (eg. humans for animals and wheat for plants) in order to investigate key traits such as development or pathology. This approach is referred to as translational biology.

How can palaeogenomics impact areas beyond fundamental science, such as agriculture and medicine?

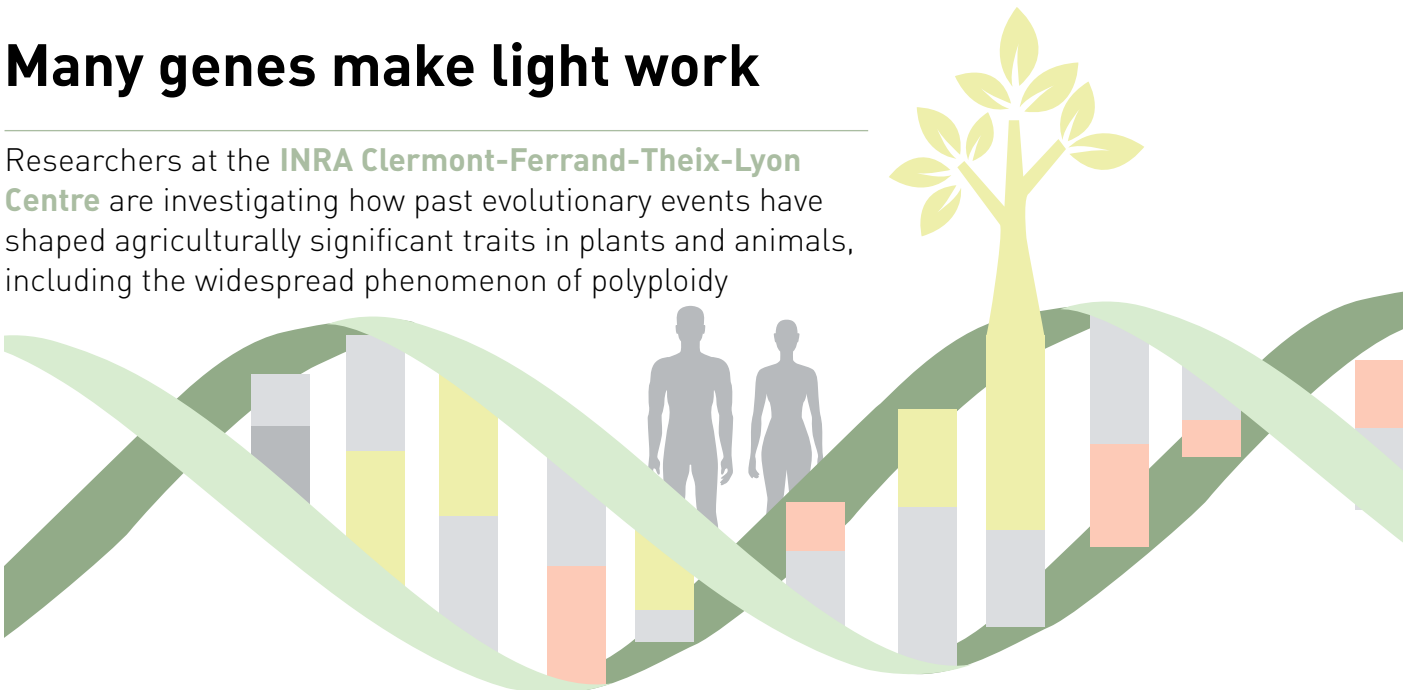
Access to ancestral genomes allows us to identify the genes that are also found in modern species. For example, a cereal ancestral genome has over 17,000 genes, shared by rice, wheat, sorghum and maize. Based on knowledge obtained from rice, it was possible to identify the genes driving key agronomic traits (enhanced yield and flour quality) in wheat. In the same way, an ancestral human genome with over 14,000 genes shared between mouse, chimpanzee and human enabled us to identify key human genes underlying diverse pathologies, based on murine experiments.

In what direction do you see your research developing in the coming years?

The ongoing deluge of genomics data requires the development of new bioinformatics tools to acquire an integrative and comparative view for fundamental and applied purposes, and ultimately to unravel key biological processes for applied biotechnology in agriculture and medicine. Palaeogenomics is entering a new era where the genomes of all living organisms, as well as DNA from fossils, will soon be sequenced to gain novel insights into the past in terms of evolution, domestication and selection. In turn, we will be better able to predict the future.

Many genes make light work

Researchers at the **INRA Clermont-Ferrand-Theix-Lyon Centre** are investigating how past evolutionary events have shaped agriculturally significant traits in plants and animals, including the widespread phenomenon of polyploidy



BACK IN 2010, a ‘draft sequence’ of the Neanderthal genome was published using material gathered from three individuals that in total provided over 4 billion nucleotides. This genetic goldmine not only demonstrated that this species is the closest known relative of *Homo sapiens*, it also helped to home in on when and where *Homo sapiens* first appeared – between 130,000 and 250,000 years ago in Africa.

Advances in extraction and characterisation techniques have made it possible to isolate DNA from preserved biological remains. Since pioneering work in 2008 comparing modern grass genomes, ancestral genomes have also been reconstructed *in silico* using mathematical methods, made possible by advances in bioinformatics. These developments have given birth to an entirely new field – palaeogenomics. By comparing DNA sequences, and even complete genomes, from ancient remains to those of modern species *in silico*, it is possible to understand the evolutionary processes that have shaped organisms alive today, including our own species.

In addition to these fascinating insights on species origin, this field can also contribute direct benefits to society – helping scientists to determine which genes underlie important traits in crops or cause disease in humans. With extensive expertise in the study of ancient genomes, Dr Jérôme Salse from the French National Institute for Agricultural Research (INRA) Clermont-Ferrand-Theix-Lyon Centre is one such scientist. Salse leads the Paleogenomics & Evolution (PaleoEVO) group, which aims to characterise the mechanisms that have shaped plant and animal genomes throughout evolution, and pinpoint the genes underlying agriculturally significant traits.

EVOLUTIONARY MODES

Palaeogenomics analyses use mathematical reconstructions of ancestral genome sequences or the sequencing of ancient

DNA to study the evolution of living species. In doing so, it can elucidate the biological mechanisms that have shaped changes to genomes in the past 500 million years of evolution. These methods have revealed that modern plant and animal lineages emerged from ancestral genomes containing just seven to 12 chromosomes and 15,000-20,000 genes. From these genes alone, the diversity of life has developed.

By reconstructing the evolutionary history of these genomes, Salse aims to unravel the similarities and differences in their structure and function. There is a great deal of variety between the two kingdoms of plants and animals, for example, with animal genomes proving relatively stable and plant genomes tending to be much more dynamic. Furthermore, their genomes differ at a number of additional levels, including chromosome number, genome size and gene number – surprisingly, perhaps, wheat possesses a genome that is five times larger than that of humans. Crucially, research done by the likes of Salse and his PaleoEVO team has helped to explain these differences.

EXPLAINING ADAPTATION

Palaeogenomics has revealed a major evolutionary mechanism for diversification: gene duplication. This process can occur in two forms – whole genome doubling (polyploidy) or local segmental duplication (small-scale duplication) – and it can have major implications for adaptation: “Having a double set of genes during evolution constitutes a route for the specialisation of duplicated gene functions to face changing environmental constraints,” Salse explains.

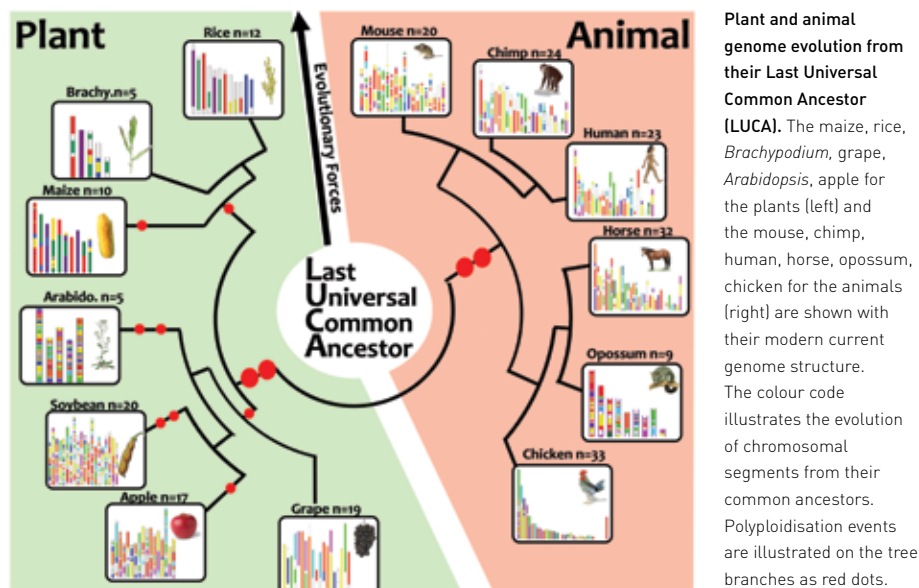
Over the course of their evolutionary history plants have made full use of this phenomenon, undergoing numerous and recurrent cycles of genome doubling over the past 250 million years. These successive polyploidisations,

which are much less frequent in animals, are believed to have triggered genetic changes leading to regular genome restructuring and reprogramming. As a result, plant genomes are much more dynamic – and faster evolving – than animals.

AGRONOMICAL INNOVATION

Salse, whose background is in plant genomics, devotes a considerable part of his work to this kingdom. His team think that these polyploidisations alter not only the structure but also the function of genes, and may explain much of the phenotypic diversity seen in plants. While the framework is in place, the details and key players in this model of evolution are yet to be provided. In characterising each step, PaleoEVO is gradually closing these gaps. The group’s efforts are divided into two key areas of investigation, beginning with the organisation of plant genomes. Access to a variety of plant genome sequences has facilitated *in silico* archaeogenomics, the reconstruction of founder genomes, which can be compared to modern genomes in order to model plant evolutionary history. These unprecedented data have revealed the evolutionary forces behind plant genomes, and shed new light on how they are organised and regulated in the present day.

The second element of the team’s research is focused specifically on the wheat genome, which is a highly useful model for understanding the impact of evolutionary structural changes. As the third most produced cereal in the world, it is also agronomically important. Wheat is grown on more land area than any other commercial food, and its world trade is greater than for all other crops combined. Using quantitative trait loci (QTL) mapping, PaleoEVO team aims to identify the genes behind important agricultural traits in wheat, such as yield. By better understanding the evolutionary basis of such traits, they aim to facilitate cloning, with the ultimate goal of



Plant and animal genome evolution from their Last Universal Common Ancestor (LUCA). The maize, rice, *Brachypodium*, grape, *Arabidopsis*, apple for the plants (left) and the mouse, chimp, human, horse, opossum, chicken for the animals (right) are shown with their modern current genome structure. The colour code illustrates the evolution of chromosomal segments from their common ancestors. Polyploidisation events are illustrated on the tree branches as red dots.

This work has helped to explain the evolutionary basis of the differences between plant and animal genomes, and will provide new tools to accelerate further progress in the field

introducing genes into existing plant breeding programmes.

TURNING THE PAGE

While the recent expansion of sequencing capacity has generated a huge amount of genomic data, it is only valuable if tools are available to handle it. To this end, in an ongoing project entitled Plant and Animal Genome Evolution (PAGE), Salse is developing novel tools to study the ancestral genomes of living organisms.

PAGE involves six pioneering research groups in the field of plant and animal evolutionary genomics from INRA, École Normale Supérieure (ENS) and the French National Centre for Scientific Research (CNRS). Together, they aim to provide – for the first time – clear insight into genome evolution by comprehensively comparing the genomes of plants and animals. It is hoped that this project will deliver fundamental knowledge as well as translational genomics-based tools to allow the scientific community to identify key traits in complex genomes by taking advantage of simpler model organisms (such as mice and *Arabidopsis*).

LOOKING AHEAD

Palaeogenomics is a powerful tool that is capable not only of providing new knowledge on how organisms have evolved, but also delivering the ability to translate knowledge from model species to more complex ones in order to investigate important traits. Salse's work has helped to explain the evolutionary basis of the differences between plant and animal genomes, and will provide new tools to accelerate further progress in the field.

Since the first ancient DNA sequence was published in 1984 from a museum specimen of an extinct subspecies of zebra, the number of palaeogenome sequences has grown rapidly. This has largely resulted from the plant and animal

POLYPOIDITY: KEY CONCEPTS

EVOLUTION

Modern plant and animal genomes are derived from ancestral genomes containing seven to 12 chromosomes and comprising 15,000 – 20,000 genes.

STRUCTURE

Genome duplications are a major mechanism of evolution. Gene content can be doubled from whole genomes (polydiploidy) or local areas (copy number variations).

FUNCTION

The gene redundancy that arises from duplications is subsequently lost through deletion or other epigenetically-driven events, subfunctionalisation (in which pairs of genes take on separate functions) or neofunctionalisation (where just one gene copy takes on a new function).

TRAITS

Genome duplications lead to genome variability, which allows organisms to adapt to new environmental constraints.

genome comparison studies that have been undertaken over the past two decades. As yet more data become available, the ability to infer evolutionary change will continue to expand, perhaps answering some of the biggest mysteries of evolution.

INTELLIGENCE

PALAEOGENOMICS: UNDERSTANDING THE PAST TO PREDICT THE FUTURE

OBJECTIVES

- To understand the evolution of modern plant and animal species by analysing ancestral genome sequences
- To transfer genes behind key biological processes from model species to species of agricultural and medical interest for applied biotechnology

FUNDING

French National Research Agency (ANR)

CONTACT

Dr Jérôme Salse

Research Director

Genetics, Diversity and Ecophysiology of Cereals (GDEC)

Paleogenomics & Evolution Research Group (PaleoEvo)
5 Chemin de Beaulieu
63100 Clermont-Ferrand
France

T +33 4 7362 4380

E jsalse@clermont.inra.fr

http://bit.ly/PaleoEvo



JÉRÔME SALSE obtained an Engineering degree in 1997 from the Graduate School of Agriculture at Toulouse (ESAP), France, before receiving his PhD from the University of Perpignan, France, in 2002. Subsequently, he also completed his Habilitation to direct research diploma (HDR) in 2009 at Blaise-Pascal University (UBP) in Clermont-Ferrand, France. Since 2003, he has been the research director of the Paleogenomics & Evolution group (PaleoEVO) at the Genetics, Diversity and Ecophysiology of Cereals research unit in the INRA Centre in Clermont-Ferrand (University of Blaise Pascal, France).