Population Genomics in Vienna, Austria

Swimming in the Pool of Genome Variation

Photo: Fotolia/Jamie Wilson

DNA sequencing is becoming ever cheaper and faster but the study of entire genomes of individuals for population-scale studies is still costly. This applies to both time and money. Christian Schlötterer and his team in Vienna have, therefore, opted to develop software tools to facilitate analysis of pooled data.

Life has colonised almost every corner on Earth – from the highest mountain to the deepest sea. It goes without saying that each location features special environmental characteristics. For example, the deep sea is always buried in pitch darkness, organisms living there are constantly 'under pressure' and even though temperatures are stable, it's still rather chilly down there (about 0°C). So, the question is how do organisms adapt to different environments and changing conditions? For Christian Schlötterer, professor at the Institute of Population Genetics in Vienna, this is the research question of his life.

Being proactive

His entire academic career has been dedicated to its investigation. "I started very early doing this. I basically scanned genomes with microsatellites [2-4 nucleotide-long repeats used as molecular markers], trying to understand how you can track genomic changes of adaptation of natural populations. I wanted to know how many genes are involved and what kind of changes are needed," he describes.

But studying genome variation among individuals of a population can translate into an expensive and time-consuming job. Even today, with sophisticated machinery on offer, sequencing individual by individual is not a feasible approach for every lab. This was even less so two years ago, when Schlötterer and his colleagues started using the Illumina sequencing technology. "At that time, this was still considerably more expensive and we were asking ourselves how we could monitor small gene frequency changes, either in experimental evolution or comparing different populations," he explains.

However, instead of tentatively sequencing through every single individual, an alternative way of tackling the population genomics issue is by working on pools of DNA from many individuals. This is what Schlötterer also had in mind but there was one problem. Software tools were not easily found for this kind of analysis. But, as we all know, necessity is the mother of invention and so the Vienna geneticists simply started to develop their own software and statistical theory.

Hands-on the code

The result was PoPoolation, a software toolbox designed for analysing sequencing data from pooled samples of a population. What the programme offers is a description of the variation among the individuals on a genome-wide scale (*PLoS One* 2011, 6(1): e15925). Robert Kofler, postdoc in Schlötterer's lab and the bioinformatician behind this software, explains that the programme is particularly helpful for identifying signatures of selection in a population. For instance, if the programme identifies unbalanced allele frequencies or, in other words, a few SNPs [single-nucleotide polymorphism] at very low frequencies, it means that only a few individuals possess those sequence variations. "This is a sign that there has been a selective sweep, showing that an advantageous mutation occurred and fixed in the population," Kofler argues.

PoPoolation reloaded

But this first effort was only useful to describe the variation in a single population. "It didn't go to the point that I actually wanted: to compare multiple populations," clarifies Schlötterer. So, not even one year after Schlötterer and co. published

PoPoolation, their upgraded tool came out: PoPoolation2 (*Bioinformatics* 2011, 27(24): 3435-6). This version is able to determine how two or more different populations compare on a genomic scale.

Kofler explains that PoPoolation2 can be used for a range of applications. One of them is selective genotyping. For example, if we have two populations of flies, one of them with black eyes and the other with white ones, we can just separate them and sequence each as a pool. Then, it is possible to compare the allele frequencies of the SNPs and determine if there are significant differences between the two groups. If we make a find, we can then identify the alleles responsible for the eye colour distinction.

Tweaking for more

Schlötterer explains that the programme can also be applied for experimental evolution studies, where you observe populations adapting to a novel environment. It is possible to measure and compare allele frequencies in this evolving population throughout the experiment at different time points. Subsequently, the experimenter can ask whether there is a specific genomic area with changes in allele frequencies. With some tweaking, it can be further adapted for other purposes.

Additionally, other flavours of this software have been developed in Schlötterer's lab. One is a database, PoPoolation DB, which can describe the basic level of polymorphism in a natural population on a genome-wide scale (*BMC Genetics* 2011, 12:27). People can go there, ask for the genome of interest and get all the variation that is present in the population. The other, and more recent one, is PoPoolation TE, where the user can estimate

the number of transposable elements and their frequencies, a tool particularly valuable for researchers studying these notorious "jumping genes" (*PLoS Genetics* 2012, 8(1): e1002487).

A software for everyone

All the programmes created by Schlötterer and his team can be easily downloaded. For instance, PoPoolation2 is freely available on *http://code.google.com/p/popoolation2*, with a well-documented tutorial and manual. This is not only helpful for the population genetics community interested in what this tool has to offer but it is also advantageous for improving the software through feedback and suggestions from the users.

Kofler, as the developer, considers that this software package is user-friendly, although he accepts that this does not guarantee that all biologists will agree with him. He explains that it is not a graphical user interface, adding that "this may be a problem

nical nature. The software was initially developed for Illumina reads and, so far, hasn't been tested with other sequencing technologies. However, Schlötterer believes that a minor change in the code could adapt to any other methodology. Moreover, he adds that Illumina is the major tool used for this kind of studies, so this might not be a huge issue.

Evolving challenges in Vienna

Kofler enjoys working with Schlötterer and doing this in Vienna. He describes the lab as having quite a stimulating environment and appreciates that they have a great amount of Illumina data. "Austria is economically in good shape but, scientifically, not the front-runner on a global scale. Having this much data is not a common thing in Austria," he declares. Like Schlötterer, he is interested in developing novel technology for analysing this enormous amount of information.

Both agree that Vienna has turned into a hub of evolutionary biology and popula-

A pool of population geneticists grouped around Christian Schlötterer (centre, light blue shirt). Robert Kofler, fourth from right, second row.

for those who are used to click buttons". But this does not mean that it is difficult to use. The user only needs to be able to handle the command line, which is not complicated and tutorials make it even simpler.

Any limitations?

Schlötterer believes that the limitations rather come from the method itself. He explains that the weakness in analysing pooled data is that "basically every read comes from a different chromosome", so it's then not possible to know whether two identified alleles come from the same or different homologous chromosomes. However, the only alternative is sequencing individuals, which is still not realistic for experimental evolution studies with multiple populations at multiple time points.

Aside from all the valuable applications, there's one minor drawback to the programme, which is, however, of a tech-

tion genetics. This is reflected by an initiative called evolVienna (*www.univie.ac.at/evolvienna*), which is intended as a communication platform among researchers established in Vienna and interested in evolution. "It also provides a forum to the outside to see how strong evolutionary biology is here. This is really impressive. I think Vienna is one of the places to be to do evolutionary science," affirms Schlötterer.

Meanwhile, in his lab, there are always new challenges. Kofler declares that all of them are very difficult to foresee and, as it happens in science, new problems arrive as old questions are being answered. The only permanent challenge for Schlötterer and his team is to deal with the great amount of data produced at the lab. How they do it, therefore, depends on technology and creativity, ingredients that are definitely not absent in this lab in Vienna.

ALEJANDRA MANJARREZ