Estimating diversity, the easy way

T Petteri Arstila

A new twist to a classic technique offers a new way to accurately measure population diversity at the nucleic acid level.

Whether the goal is to find out just how homogeneous a commercially important domesticated plant is, or how many variants a virus has produced during its years of persistence in a patient, estimating diversity is important in many fields of research. Unfortunately, diversity analysis often entails laborious cloning and sequencing, making the analysis of more than a handful of samples unfeasible and imposing limits on what can be directly measured. In this issue, Baum and McCune describe a method that might make things considerably easier in the future[1]. Within hours, using only a real-time PCR instrument, the method provides an accurate and reproducible estimate of the number of sequence variants in a sample.

The new work gives a clever twist to an old method called concentration-time, or C_0t analysis, which had been used to estimate genome complexity before rapid sequencing became available[2]. Because the starting material is first amplified by PCR, the authors call their new assay AmpliCot. The basic principle is simple (**Fig. 1**). The amplified DNA is mixed with SYBR Green, a fluorescent dye that binds double-stranded DNA. Heating the sample to 95 °C denatures the DNA completely, and the fluorescence is lost. The temperature is then decreased enough to allow reannealing. As more and more homoduplexes accumulate, the fluorescent signal increases again until its intensity reaches that of a nonmelted reference sample.

The essence of the AmpliCot is in the kinetics: the more sequence variants there are in the sample, the slower the reannealing will be. The time needed to reach a given fluorescence is called the 'Cot value' and, when compared with a sample of known diversity, allows the estimation of the number of different sequences in the sample. The target sequences must contain conserved sequences for PCR primer annealing at both ends, and the variable part must contain enough nucleotide differences to allow reliable differentiation between heteroduplexes and homoduplexes on the basis of melting temperature. Within these constraints, any kind of sample can be analyzed.

The application chosen by the authors as a proof-of-principle was to measure the diversity of the T-cell receptor (TCR). The TCR is used by T cells to recognize antigens, and is therefore the key controller of T-cell activation and adaptive immunity[3]. During T-cell development, recombination of the gene segments encoding the TCR has the potential to generate a practically infinite repertoire of different, clonally distributed TCR sequences, each of them expressed by

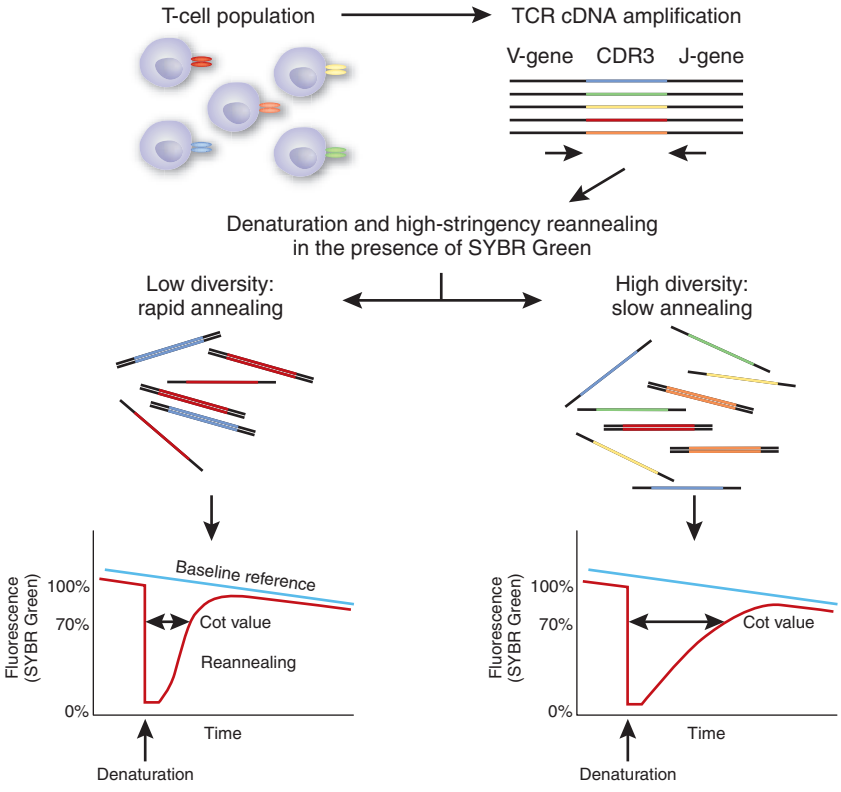

Figure 1 | TCR repertoire diversity of a T-cell population analyzed by AmpliCot. After amplification of the TCR hypervariable region, CDR3, the PCR products are denatured and reannealed in the presence of the dsDNA-binding dye SYBR Green. The time needed for DNA reannealing, the Cot value, is proportional to the number of different sequence variants in the sample. A nondenatured reference sample is used as baseline reference to control for fluorescence decay.

T. Petteri Arstila is at the Haartman Institute, Department of Immunology, PB21, 00014 University of Helsinki, Finland.
e-mail: petteri.arstila@helsinki.fi

a tiny subset of T cells[4]. With such diversity, studying the behavior of the vast TCR repertoire at sufficient resolution is a daunting task, usually requiring extensive sequencing, and even then only a fraction of the repertoire can be directly analyzed.

Most of the variation resides in the junctional region between the rearranged TCR gene segments, the complementarity determining region 3 (CDR3), which can thus be used as a tag for the entire gene encoding the TCR. Conveniently, the hypervariable CDR3 is flanked by conserved sequences, creating the necessary conditions for AmpliCot analysis. As the authors carefully demonstrate, their assay gives results very close to those previously obtained by painstaking sequencing.

In diverse populations, including T cells, the variants are usually present at very different frequencies, so any method used to estimate the diversity must also be able to detect rare clones. In AmpliCot, abundant sequences account for most of the early reannealing, so to maximize the detection of rare variants the Cot values should be taken at the late phases of the reannealing curve. Baum and McCune showed that by doing so, they were able to accurately measure the full diversity of their samples even in the presence of dominant sequences.

The upper limit of AmpliCot's resolution is unknown; the highest TCR diversity actually measured by the authors is roughly 3,500 different sequences. Another point to note is that AmpliCot does not provide any sequence data, or information about the clonal composition of the sample. In many settings obtaining such information is necessary; rather than replacing the traditional methods, in these cases AmpliCot will provide a complementary approach.

Nevertheless, the speed and simplicity of AmpliCot will open entirely new possibilities to analyze the kinds of questions that have been very difficult to address by sequencing. For example, it should be easy to monitor, with repeated samples, the evolution of an antiviral T-cell response, and at the same time measure the genetic drift of the pathogen. The mauling given by HIV to the T-cell repertoire could be analyzed, as well as the subsequent effects of antiretroviral treatment. Another realm of application might be in the area of bone marrow transplantation, as the prognosis of transplant recipients is likely to be affected by the speed with which they regain full T- and B-cell diversity. There also, AmpliCot may offer a clinical tool to monitor these patients.

And the potential applications are certainly not limited to the study of the immune system. From human genetics to microbiology, and many other fields besides, AmpliCot is likely to find innovative uses.

1. Baum, P.D. & McCune, J.M. *Nat. Methods* **3**, 895–901 (2006).
2. Britten, R.J. & Kohne, D.E. *Science* **161**, 529–540 (1968).
3. Davis, M.M. *et al. Annu. Rev. Immunol.* **16**, 523–544 (1998).
4. Nikolich-Zugich, J., Slifka, M.K. & Messaoudi, I. *Nat. Rev. Immunol.* **4**, 123–132 (2004).