An alignment editing tool for the design of nucleotide sequence-based assays

Piotr Wojtek Dabrowski[2], Michael Goltz[2], Andreas Nitsche[1]
1 Robert Koch Institute, Centre for Biological Security 1, Berlin, Germany
2 Robert Koch Institute, IT Department

Diagnostics and quantification of viral pathogens is usually based on the detection of viral nucleic acids via real-time PCR amplification (Heid, Stevens et al. 1996). With carefully designed probes, this approach also often allows the identification of viral species based on characteristic single nucleotide polymorphisms (SNPs) using methods such as melting analysis (Herrmann, Durtschi et al. 2006). Alternatively, the PCR product can be sequenced to obtain information on such polymorphisms. Unfortunately, there are cases where no suitable stretches of sequence can be found that contain SNPs characteristic for all species of interest. To circumvent this problem, multiplex pyrosequencing can be used, wherein several PCR products with several pyrosequencing primers are sequenced simultaneously (Patel, Shiao et al. 2007).
The design of nucleotide sequence-based assays – PCR, real-time PCR or multiplex pyrosequencing assays – can be a challenging task. Especially when primers and probes for the differentiation of species are required and a high number of sequences is known, the amount of data which needs to be taken into account can be overwhelming. Therefore, we have developed a sequence editor which implements several features specifically geared towards assay design, none of which are available in other sequence editors to the best of our knowledge. Firstly, sequences within an alignment can be (manually or automatically) grouped, and consensus sequences can be calculated both for the alignment globally and for each group individually. Groups can be collapsed, leaving only the conservation graph and the consensus sequence visible. This allows the user to work with a much reduced amount of data while still retaining access to all relevant information. SNPs which are conserved within a single group and are thus candidates for use in a differentiation assay can be automatically detected and highlighted. Some basic primer design functionality (such as Tm calculation, product size calculation, degeneration etc.) is supported. Novel functionality is provided for the design of multiplex pyrosequencing assays: Potential sequencing primers can be marked as pyrosequencing primers and the pyrograms which would be generated using these primers in a multiplex pyrosequencing assay can be displayed. If the expected pyrogram is not unique for each of the defined groups, a warning is displayed (see Figure 1).

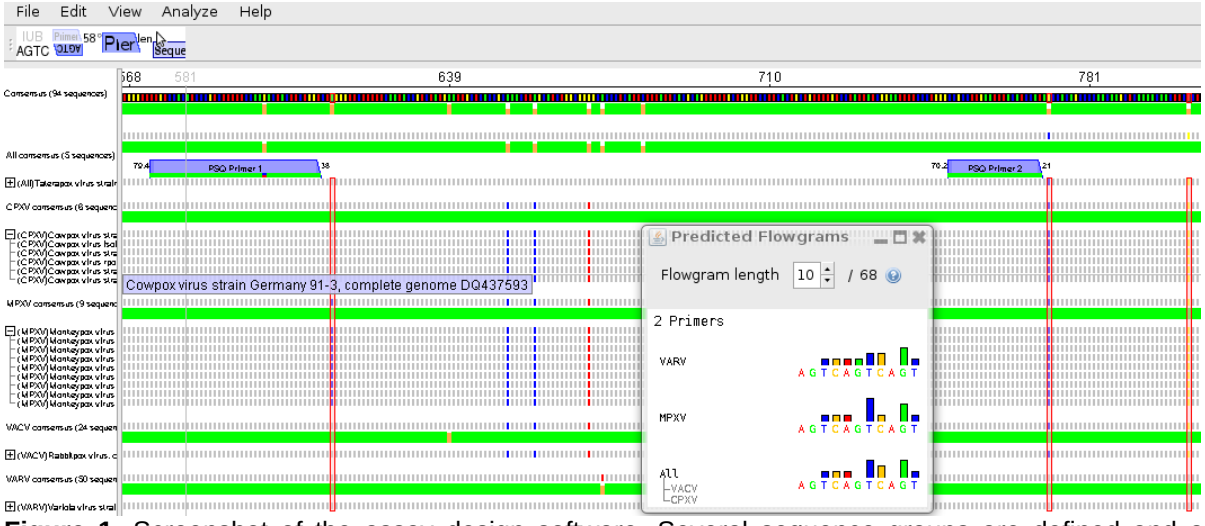

Figure 1: Screenshot of the assay design software. Several sequence groups are defined and a consensus sequence including conservation is shown for each group. The two groups at the bottom are collapsed. Conserved SNPs are highlighted in red. Two pyrosequencing primers (blue annotations) have been defined and the predicted pyrograms for each group are shown. As can be easily seen the pyrograms for the groups VACV and CPXV are identical, which shows that this exemplary assay cannot be used to differentiate between these two viral species.

References

Heid, C. A., J. Stevens, et al. (1996). "Real time quantitative PCR." Genome Res. **6**(10): 986-994.

Herrmann, M. G., J. D. Durtschi, et al. (2006). "Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes." Clin Chem. **3**(52): 494-503.

Patel, P., Y. H. Shiao, et al. (2007). "Multiplex pyrosequencing for DNA variation analysis." Methods Mol Biol.(373): 75-88.