Supplementary information

ReX – A suite of online tools for the design, visualization and analysis of chimeric protein libraries

Weiliang Huang[1,2*], Wayne A. Johnston[2], Mikael Boden[2] and Elizabeth M.J. Gillam[2*]

Demonstration: choosing REs for DNA shuffling of members of the cytochrome P450 2A subfamily using REcut

To demonstrate of the use of REcut, DNA sequences of cytochrome P450 2A (CYP2A) subfamily members, CYP2A5, CYP2A6 and CYP2A13 were loaded by pasting their coding sequences in FASTA format in the web interface. The settings of maximum fragment length and minimum fragment length were 400 bp and 50 bp, respectively. The extensive search function was enabled to allow evaluation of up to triple-enzyme combinations of REs. Thus, more useful combinations were revealed at the expense of computation time. After the analysis completed, the sets of REs that could digest all parents into DNA fragments within the size limits and the corresponding digestion patterns were displayed. Parents were displayed in the cleavage patterns in the same order as provided in the input. An example view is provided in Supplementary Figure 1.

RE-mediated DNA shuffling requires two sets of REs to generate two sets of DNA fragments with overlapping ends to allow ORF reassembly by PCR. Thus, fully or partially overlapping cleavage sites should be avoided, since the sequence left for priming would be usually too short to allow effective annealing. Considering that the CYP2A parents share 80% parent sequence identity and an average GC content of

50%, self-priming of two overlapping fragments requires the two ends to hybridize with at least 19 bases overlap to ensure proper annealing at a T_{ann} above 45 °C, estimated using the formula of Marmur (1). In REcut, the option of automatic selection on the preview page shown in Supplementary Figure 1 was chosen with the minimum bases required for annealing set to 19 bases. REcut checked the priming compatibility between DNA fragments generated by different RE sets and selected suitable combinations automatically. An example of the output provided is shown in Supplementary Figure 2; RE sets appropriate for generating fragments that can anneal efficiently in a reassembly PCR are listed.

Alternatively, by selecting the manual option as shown on the preview page depicted in Supplementary Figure 1, users can also choose specific RE sets and let REcut verify their complementarity. An example view of digestion patterns generated by two sets of manually selected REs is provided in Supplementary Figure 3.

Application of REcut to the fragmentation of three CYP2A sequences for a DNA shuffling library produced a mean fragment size ideal for reassembly around 217 bp without extremely small or large fragments and also with a low size variation (CV=0.36).

Figure 1. The preview result page of REcut showing cleavage patterns of REs sets that can cleave all parents into DNA fragments within the size limit set by the user (*only part of the output is shown here to save space*). Parents are displayed in the

order of entry. REcut not only offers the option to further screen RE sets that can generate DNA fragments able to re-assemble automatically, but also allows the user to select RE sets manually and will analyze the PCR priming compatibility of the user-specified selection.

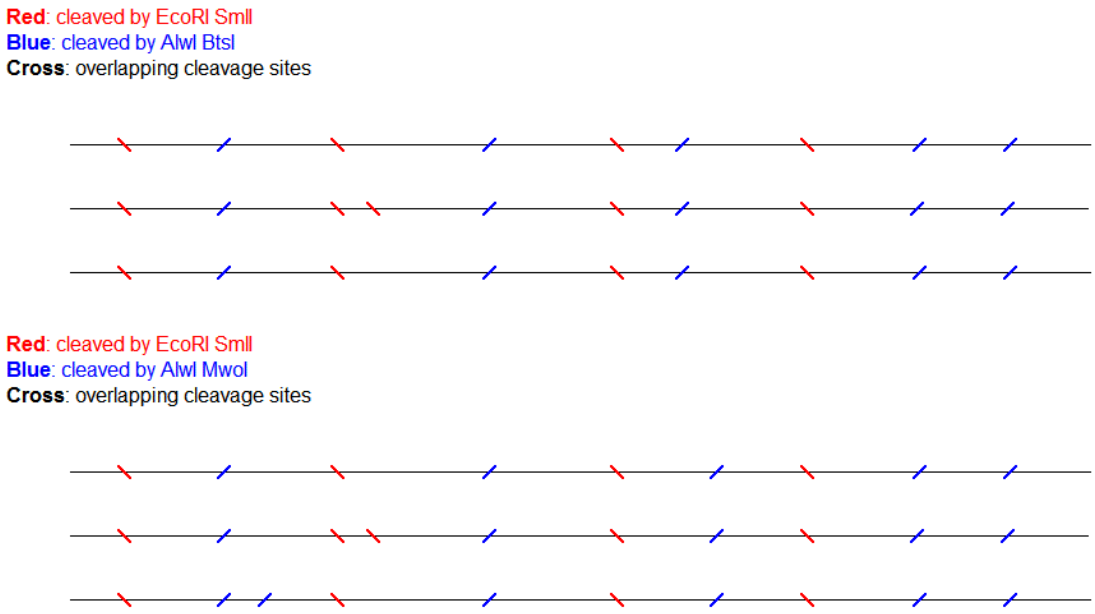

Figure 2. An example view of alternative restriction digestions chosen automatically by REcut, which can cleave parents into DNA fragments that can efficiently reassemble during PCR. Each RE-mediated DNA shuffling experiment requires two sets of REs, whose cleavage sites are drawn in red and blue respectively. As shown in the figure, there are no overlapping cleavage sites (i.e. sites are all separated by at least 19 bp), which, if present, would be shown as red-blue crosses. The user can assemble DNA fragments generated by RE sets from any of these combinations displayed.

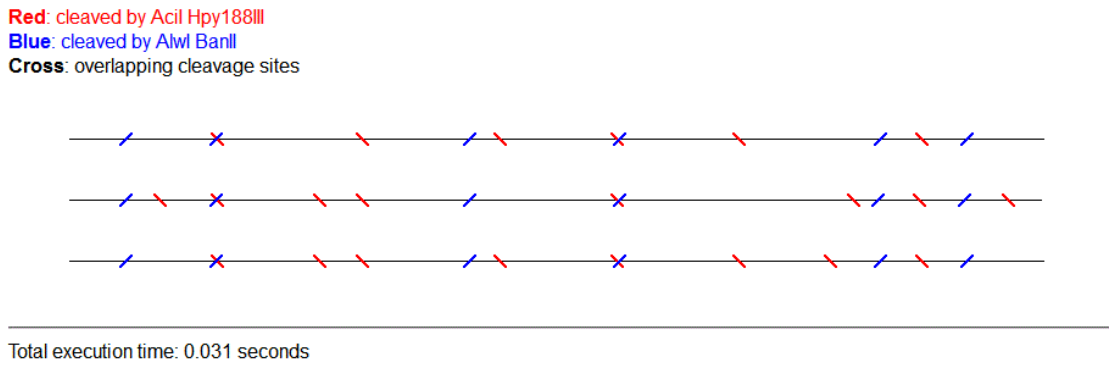

Figure 3. The self-priming verification performed by REcut on two RE sets manually selected by the user revealed that the first RE set of *BstY*I, *Cla*I *and Rsa*I has overlapping cleavage sites with the second RE set of *BtgZ*I, *EcoN*I *and Tsp509*I on two parents, marked by red-blue crosses, which should be avoided for DNA shuffling.

1.**Marmur, J. and P. Doty.** 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J Mol Biol 5:109-118.