**Detection of potential chimeric sequences in HOMD 13.2 and HOMD ext. 1.1**

The following was performed, using MOTHUR 1.33, to detect potential chimeras in both sets:

**Chimeras in HOMD 13.2**

1. Full 16S rRNA sequences of 21 novel oral taxa recently described by Camanocha et al. were added.to HOMD 13.2.
2. Sequences were then aligned to SILVA.
3. Sequences of **named species** in HOMD 13.2 were extracted and added to those of SILVA gold to make a reference set\* after removing redundant sequences.
4. Both Uchime and Chimera Slayer were used to detect chimeras using the combined reference set.
5. Chimeras were removed (updated-HOMD 13.2) †

**Chimeras in HOMD ext 1.1**

1. HOMD ext 1.1 sequences were aligned to SILVA
2. **All** updated-HOMD 13.2 sequences were added to those of SILVA gold to make a reference set \*
3. Both Uchime and Chimera Slayer were used to detect chimeras
4. Chimeras were removed (trusted-HOMDext)†

\* SILVA gold is recommend for use as a reference with chimera detection software. However, since almost all sequences in the samples belong to oral taxa, HOMD 13.2 sequences are theoretically the best reference for detection of chimeric sequences so were added to SILVA gold. In fact, one established method for detection of chimeras in NGS data is using high abundance sequences in the dataset itself; compared to this, using HOMD 13.2 sequences is even more reliable. More chimeras were identified in the NGS dataset using the combined SILVA gold and HOMD 13.2 reference set than with using SILVA gold alone.

† Among the detected chimeras were cultivated phylotypes. While these probably represent false positives, they were kept away for re-assessment of their taxonomy and sequence quality.