Preassembled genomes are the most accessible source of defined genomic segments, as the problems of stitching together overlapping sequence fragments have already been tackled and the assemblies will have been subject to some degree of validation and quality control. Complete assemblies can be obtained from a number of disparate sites depending on the organism and assembly method. However, the UCSC Genome Browser (http://genome.ucsc.edu/), Ensembl (http://ensembl.org/) and the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) all provide portals to the most current, and archived public assemblies. These sites also provide means of searching the assemblies, such as BLAST (Altschul *et al.*, 1997), BLAT (Kent, 2002) and SSAHA (Ning *et al.*, 2001) as well as precomputed annotation for the genome assemblies that can be readily incorporated into comparative genomic analyses.

There are several routes to identifying homologous loci in target genome sequences. An obvious approach is based on sequence similarity searches, but care must be taken to distinguish orthologous from paralogous loci. Processed pseudogenes, in particular, are common (Shemesh *et al.*, 2006); these are the reverse-transcribed copy of an mRNA that has integrated into the genome, but which does not code for a functional protein (Section 6.2). As processed pseudogenes lack introns, they can score better than an orthologous locus in a similarity search. Genome-wide, reciprocal best matches (Tatusov *et al.*, 2003) can be used to increase confidence that two loci are orthologous. Ensembl also provides precomputed assignments of gene orthology, currently based on reciprocal best matches for several genomes in the '*geneview*' pages and from the *EnsMart* data repository. Conservation of the order and orientation of genes in and neighbouring the locus can also provide additional support of the orthology of two loci.

Probably the simplest currently available route to identifying orthologous loci is with the *Net* alignments at UCSC. These genome-to-genome pairwise alignments show genome-wide best matches and local rearrangements within them. They provide a direct means of jumping between an orthologous location in two genomes and can be used directly to delineate an orthologous locus in a target genome. For example, with the genome browser showing a complete locus of interest in a human assembly, clicking on the human to dog *Net* will provide an option to open the dog genome browser in a corresponding window, from which the canine sequence and associated annotation can be obtained. An extension of this method is to use the genomic alignments to transfer annotation from one perhaps well-annotated genome to another that may have been recently assembled. The LiftOver tool at the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgLiftOver) provides this facility for a limited set of genome pairs. This can provide a rapid way to get a baseline annotation, which can then be filtered and refined. The *Net* alignments are generally good quality, but problems do arise, in particular where segmental duplications and assembly gaps are involved.

If there is uncertainty in the assignment of paralogy or orthology between multiple sequences, it can often be resolved through rigorous phylogenetic analysis, of