-examine a very broad phylogenetic context (animals vs plants)

-an orthoganol way to investigate homology when phylogenetics is uncertain

-synteny data can speed the transfer of knowledge from model to non-model systems

-differences in genomic organization between plants and animals due to1) differences in fundamental molecular processes (eg DNA repair + recombination) and historical biology (mode of reproduction, generation time, pop sizes)

Their approach: Fig. 1

1. Perform n2 pairwise reciprocal all vs. all comparisons of all annotated genes in all species (i.e. eg. all vs all blasts) = gene similarity relationships.
2. For each species pair, input gene similarity relationships + gene positions (GFF) into a synteny analysis program (eg. MCScanX) for collinearity/microsynteny block detection. Store as a n(n+1)/2 (i.e. half diagonal square) matrix (species x... vs species y...)
3. Traditionally, related blocks centred on a target locus (aka microsynteny block families) are organized into parallel coordinate plots. This paper organizes them as networks via Infomap (a k-clique percolation method?): genes are nodes and edges between them represent a syntenic relationship across chromosomes of different animals.
4. ? The entire network database contains phylogenetic syntenyt trajectoris of all annotated genes which can be used for further analyses

Genomes vary in quality (# scaffolds, N50, BUSCO). Test MCscanx settings (see p. 2168)

Fig. 2 Heatmap of pairwise collinearity/microsynteny comparisons: Broad patterns in synteny across long evolutionary distances

-Diagonal (self vs self): Intragenome comparisons - reflect orthologs and Ohnolog (gene duplicate arising from a WGD). We are unsure why they are in blue – should be 100.

-can pick out chunks

Comparison with quality scores: syntenic percentage positively correlates with N50 and BUSCO and **neg correlates with genome size and number of scaffolds >> relate to your own findings. How would your approach alter this relationship, if at all?**

Fig. 4B: clustering statistics

-For mammals, node degree peaks ~70-80, where cumulative fraction of nodes peak from 0.2 to 1

-there are 70-80 high quality mammal genomes, thus node degree matches this number because all these genomes are high quality enough to detect synteny

-peaking from 0.2 – 1 = most nodes have same number of links

-clusters with more genes could represent several rounds of WGD or tandem-dupicated ararays eg hox genes. Small clusters might be lineage-specific transpositions – synteny shared only across a few closely related species

-see other stats: glossed over

Fig. 5: heatmap of genes vs species, with color indicate number of gene copies in each synteny block

-slightly confused whether eg. “single-copy syntenic clusters” - i.e. syntenic clusters contain only genes that are single copy, whereas duplicate-copy clusters are syntenic blocks, each with a duplication in them, that are syntenic with another animal?

-see fig. legend

Fig. 6

-if BUSCOs are in different synteny clusters, they were found in different synteny family blocks (i.e. > 1 location in genome)

-so if in only 1 cluster = BUSCO gene is syntenic across all species

Rebel genes: from BUSCO gene data

-in mammals, 2 or more synteny clusters are less common

-in plants, single synteny clusters are less common

-characterized with GO enrichment analysis