Best Evidence Sources

MCOT original
(from FASTA files)
genome-independent assembly of RNA-seq reads
usually very accurate sequence (Illumina)
special selection process to choose most complete transcript available
still may not be full-length

IsoSeq HQ original
(from citrusgreening.org BLAST)
produced by PacBio sequence technology
best source for gene structure because they are genome-independent and no assembly was needed
often full-length because of long read sequencing technology
sequence supposed to be 99% accurate but could have small indels

Isoseq-hq mapped
(track in Apollo Diaci_v3.0)
When available, this is a good model to start with, but genome assembly errors could affect accuracy of mapped model, so compare to IsoSeq HQ original sequence. Significant differences likely indicate a problem with the assembled genome.
However, if there are SNPs between the mapped sequence and original IsoSeq sequence, the mapped sequence may actually be more accurate (Illumina vs PacBio sequence).

Good Evidence Sources

MCOT mapped
(track in Apollo Diaci_v3.0)
Also usually a good model to start with, but needs to be compared to original MCOT sequence, because it could be affected by genome assembly errors and/or indels causing frameshifts

de novo transcriptome - original
(from Dcitri_transcriptome.fasta file)
made up of all IsoSeq transcripts and transcripts assembled from all available RNA-seq data
genome-independent source
lots of partial transcripts
low quality IsoSeq transcripts have not been removed (these can have fusions and other errors)

de novo transcriptome - mapped
(track in Apollo Diaci_v3.0)
could be affected by genome assembly errors and/or indels causing frameshifts

low quality IsoSeq transcripts have not been removed
Mapped IsoSeq transcripts often have more accurate sequence than the original IsoSeq reads.

RNAseq/Mapped Reads

(multiple tracks in Apollo Diaci_v3.0)
Great for determining intron/exon boundaries
Can also be used as a genome-independent sequence source in case of sequence discrepancies

RNAseq/Quantitative tracks

(multiple tracks in Apollo Diaci_v3.0)
Good for quickly checking whether intron/exon structure of model is supported by RNAseq data
Usually all exons in one gene will show similar expression levels - could help identify fusion models
Expression levels at each end of gene may be slightly lower due to degradation of cDNA ends.

Fair Evidence Sources

OGS3.0 gene models final beta

(track in Apollo Diaci_v3.0)
Computationally predicted models based on genome v3.0 + manual annotations from all versions so far
Used complete MAKER pipeline
Accurate models for maybe 50 percent of genes
Has quite a few models that mistakenly fuse adjacent genes into one model
Most manually annotated models have the human-given gene name as the description
Computationally predicted models have AHRD annotations as description

Less Reliable Evidence Sources

OGS1 curated genes - original

(from dcitr_manualcuration_12-22-2016_trans.fa)
These were curated by humans, so usually better than computationally predicted models from the same genome.
However, they are based on genome v1.0, so they may be partial or contain various errors

OGS1 curated genes - mapped

(track in Apollo Diaci_v3.0)
Differences between genome v1.0 and genome v3.0 may cause mapping difficulties.
Many models are partial or have other errors because of problems with the v1.0 genome assembly

OGS2.0 mRNA mapped

(track in Apollo Diaci_v3.0)
computationally predicted models based on genome v2.0

genome had lots of assembly errors (especially duplications), so models may have duplicated exons.
Many partial models
Some models incorrectly fuse adjacent genes

x7 no SNAP gene models
(track in Apollo Diaci_v3.0)
computationally predicted models based on genome v2.0
Left out SNAP models from MAKER pipeline to get rid of fusion gene artifacts
Fewer fusion gene artifacts, but also fewer accurate models

NCBI genes mapped
(track in Apollo Diaci_v3.0)
computationally predicted models based on genome v1.0
Lots of errors, not very reliable