Published in final edited form as:

Plant Biotechnol J. 2013 September ; 11(7): 883–893. doi:10.1111/pbi.12081.

Comparative Genomics Analysis in Prunoideae to Identify Biologically Relevant Polymorphisms

Tyson Koepke^{1,2,*}, Scott Schaeffer^{1,2,*}, Artemus Harper¹, Federico Dicenta³, Mark Edwards⁴, Robert J. Henry⁵, Birger Lindberg Møller⁶, Lee Meisel⁷, Nnadozie Oraguzie⁸, Herman Silva⁹, Raquel Sánchez-Pérez^{3,6,§}, and Amit Dhingra^{1,2,§}

¹Department of Horticulture, Washington State University, Pullman WA, USA

²Molecular Plant Sciences Graduate Program, Washington State University, Pullman, WA, USA

³Department of Plant Breeding, CEBAS-CSIC, PO BOX 164, 30100 Espinardo, Murcia, Spain

⁴Southern Cross University, Lismore NSW 2480, Australia

⁵Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia QLD 4072

⁶Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Denmark

⁷INTA-Universidad de Chile, Santiago, Chile

⁸IAREC, Department of Horticulture, Washington State University, Prosser, WA, USA

⁹Laboratorio de Genómica Funcional & Bioinformática, Departamento de Producción Agrícola, Facultad de Ciencias Agronómicas, Universidad de Chile, 8820808, La Pintana Santiago, Chile

Abstract

Prunus is an economically important genus with a wide range of physiological and biological variability. Using the peach genome as a reference, sequencing reads from four almond accessions and one sweet cherry cultivar were used for comparative analysis of these three *Prunus* species. Reference mapping enabled the identification of many biological relevant polymorphisms within the individuals. Examining the depth of the polymorphisms and the overall scaffold coverage, we identified many potentially interesting regions including hundreds of small scaffolds with no coverage from any individual. Nonsense mutations account for about 70,000 of the 13 million identified single nucleotide polymorphisms (SNPs). Blast2GO analyses on these nonsense SNPs revealed several interesting results. First, nonsense SNPs were not evenly distributed across all gene ontology terms. Specifically, in comparison to peach, sweet cherry is found to have nonsense SNPs in two 1-aminocyclopropane-1-carboxylate synthase (ACS) genes and two 1-aminocyclopropane-1-carboxylate oxidase (ACO) genes. These polymorphisms may be at the root of the non-climacteric ripening of sweet cherry. A set of candidate genes associated with bitterness in almond were identified by comparing sweet and bitter almond sequences. To the best of our

§Corresponding authors: adhingra@wsu.edu.

*The authors wish it to be known that the first two authors would like to be regarded as co-first authors.

The authors declare no conflict of interest.

Author contributions: AD and HS led the sweet cherry genome data generation as part of the sweet cherry genome consortia with contribution from NO and LM. RPS led the almond genome data generation with contributions from FD, BLM, ME and RH. TK, SS, and AD designed the analyses. TK and AH completed the reference mapping, mutation analyses and analysis of the *Sk* locus. SS performed the BLAST2GO analysis and processing. TK and SS performed statistical analyses. SS, TK, and AD wrote the first draft of the manuscript. AD and RPS supervised the study. All authors contributed to, read and approved the final manuscript.

knowledge, this is the first report in plants of nonsense SNP abundance in a genus being linked to specific GO terms.

Introduction

Genetic and genomic diversity arises through multiple mechanisms including whole genome duplication, gene copy and transposable elements. However, in closely related species, and especially within varieties, single nucleotide polymorphisms (SNPs) play a large role in contributing to genetic variation. The SNP differences in closely related species and varieties, determine the phenotypic diversity observed in these plants. While large scale rearrangements, duplications and deletions contribute to genetic changes, SNPs as well as insertions and deletions (indels) can have a direct effect on gene expression and function. SNPs and indels can be rapidly assessed through high-throughput sequencing and re-sequencing efforts and are becoming widely used as genetic markers in breeding programs (Ahmad et al., 2011; Ganai et al., 2009; Hyten et al., 2010; Kulheim et al., 2009).

While most previously identified polymorphisms have been the result of intraspecific analyses, the genetic changes contributing to the phenotypic variation across different species of a genus are also of interest. *Prunus*, a diverse genus in the Rosaceae family with economically important ornamentals, fruits, seeds, and wood based products, is a good candidate genus for this type of analysis. *Prunus* contains species that are diploid with $n=x=8$ and have estimated genome sizes between 225Mb and 300Mb (International_Peach_Genome_Initiative, 2013; Shulaev et al., 2008; Zhebentyayeva et al., 2008), relatively small for the Rosaceae family. Peach has also been established as a reference genotype for *Prunus* due to the vast genomic resources available for peach including many ESTs, DNA markers, and linkage maps (Zhebentyayeva et al., 2008). The recently completed draft genome sequence of peach is 220–230Mb (International_Peach_Genome_Initiative, 2013).

Peach, almond, and sweet cherry production was collectively valued at over 3.6 billion dollars in the US in 2010 (NASS, 2011) demonstrating the economic importance of this genus and the value of understanding the genomic structure of these species. Each of these are crops in the *Prunus* genus, that produce stone fruits, have a perennial growth habit, and have a prolonged juvenility stage that has hindered the rate of progress of conventional breeding and genetic analyses.

While these species are closely related, they have differences in economically important traits that are important to production. In almond, the primary trait of interest is the difference between bitter and sweet almonds, though flowering time and shell hardness are also important. Bitterness in almonds is driven by the production of amygdalin, a cyanogenic diglucoside and its degradation products benzaldehyde and cyanide (Sánchez-Pérez et al., 2012; Sánchez-Pérez et al., 2008). This trait has been found to be controlled by a single, dominant gene called *Sweet kernel* (*Sk*) that produces sweet almonds (Dicenta and García, 1993; Dicenta et al., 2007). SSR markers have placed *Sk* on linkage group 5 of the “R1000” and “Desmayo Langueta” almond maps (Sánchez-Pérez et al., 2010). The position of these SSRs on the almond “Texas” × peach “Earlygold” (TxE), places the *Sk* locus between 11Mb and 14.6Mb on the peach scaffold 5 (Sánchez-Pérez et al., 2010). Several targets for DNA markers in sweet cherries are: fruit size, firmness, pedicel-fruit retention force (PFRF), and powdery mildew resistance. Since the peach genome is available and intra-specific polymorphism analyses have already been concluded in peach (Ahmad et al., 2011), this work focuses on the genomic differences of almond and cherry in respect to peach.

Here, a reference mapping approach using the peach genome v1.0 as the reference genome and high-throughput sequencing from four almond accessions and one sweet cherry cultivar were used to identify regions of increased and decreased conservation in *Prunus*. Detailed analysis of SNPs and indels was completed to build a resource for future inquiries into these species. Additionally, preliminary analysis of the *Sk* locus in almond was completed, identifying 228 SNP candidates associated with the *Sk* gene. The collective polymorphism dataset provides several regions of interest that have lower polymorphism rates and may be essential to the shared characteristics of these *Prunus* species.

Results

Sequencing data acquisition

Four almond accessions were chosen for sequencing including 2 sweet cultivars, Ramillete and Lauranne, and 2 bitter selections of CEBAS-CSIC, D05-187 and S3067. Shotgun sequencing of these four almond genotypes produced 142 million 76-base Illumina reads. Each of the individual almond genotypes was sequenced at 8–13x coverage, or 2.1–3.3 Gb of sequence, and combined to yield a 10.8 Gb dataset or 43x coverage (Table 1). The sweet cherry cultivar Stella was chosen for genomic sequencing, and through 454 single-end reads, 454 paired-end reads, and Illumina paired-end reads, 1.6 Gb of sequence or roughly 7x coverage was acquired. Transcriptome sequencing of sweet cherry produced an additional 460Mb of sequence of single end 454 reads from Bing and Rainier cultivars. The raw data were submitted to NCBI SRA (accession number SRP020000).

Assembly

A reference based assembly of the reads onto the v1.0 of the peach genome (International_Peach_Genome_Initiative, 2013) was completed to identify regions of conservation and divergence in the *Prunus* genus. Out of all the combined Illumina reads 56% mapped to the peach nuclear genome. 99% of the remaining reads (44% of the total reads) mapped to the peach chloroplast genome. Only 0.2% of the total reads did not map to either. This confirmed that the mapping was efficient and the chloroplast mapped reads were not analyzed further. The eight primary scaffolds of peach were covered between 0.4 and 6.3x as shown in 'Additional File 1' which contains the coverage statistics for each scaffold and data set. These scaffolds were covered an average of 4.94x for the combined cherry data and a 3.14x average for the almond genotypes. Overall, 162 of the 334 scaffolds contained zero reads from cherry or almond while an additional 24 were not mapped by the cherry data. Also, mapping data show that 96–99% of peach genes were mapped to with these datasets (Table 2).

Polymorphism Analyses

Overall, 13,126,567 initial polymorphisms were identified between each individual genotype and peach. The raw SNP report is available from authors upon request. Potential polymorphisms were initially identified and parsed to 9,751,035 after filtering to retain only sites with at least 3 reads supporting the difference as previously described (Deschamps and Campbell, 2010; Hyten et al., 2010; Koepke et al., 2012; Kulheim et al., 2009). These polymorphisms were then further identified based on their position revealing a total 6,138,404 polymorphic sites.

Polymorphism type and region identification—Based on the reference genome annotations (Additional File 2) from GDR (Jung et al., 2008), the polymorphisms passing the filtering criteria were classified by their location (Table 3) yielding an average of 260,000 polymorphisms in the coding sequence (CDS) for the almond accessions and greater than 300,000 polymorphisms in sweet cherry. Polymorphisms in the exons of

cDNAs (Additional File 3) of the almond genotypes average 52.1% (155,010), 43.3% (128,778), and 4.5% (13,454) for sense, missense, and nonsense mutations respectively (Table 3). Additionally, 0.1% (342) of the CDS SNPs are read-through mutations, mutations modifying a stop codon into an amino acid yielding C terminus extension also termed 'stop loss mutations' (Zirn et al., 2005). Sweet 1, however, had a much higher rate of nonsense mutations (10.6%), while the other 3 genotypes had fewer nonsense mutations (2–3%). The insertions and deletions in the exons averaged ~3,000 each for the four almond genotypes. In the cherry genomic dataset, exonic SNPs were 50.4% (162,662) sense, 42.8% (137,976) missense, 6.6% (21,234) nonsense, and 0.1% (335) read-through mutations along with 16,155 indels (Table 3).

Polymorphism depth analyses—The passed filtering dataset was also used to analyze the occurrence patterns of the polymorphisms. For scaffold 1 (Additional File 4), it is clear that there are several regions of interest containing significantly more or less than the average number of polymorphic sites. Similar mapping of the number of genes in these regions of the peach scaffold reveal low gene density regions with high polymorphism rates. Statistical analyses reveal 346 sections that significantly differ from the mean number of polymorphisms in each 50kb region on each individual scaffold (Additional File 5). 95 of these sections combine to make 31 regions that are greater than 100kb in length with the longest region containing significantly higher polymorphisms being a 600kb block in almond from 20.45Mb–30Mb on scaffold 1. This region in cherry contains two 50kb blocks and one 100kb block that are also significantly higher in polymorphism rate. These genomic regions may potentially be the regions responsible for phenotypic divergence from other members of *Prunoideae*.

Analysis of *Sk* locus

Further filtering of the almond polymorphisms around the *Sk* locus was completed to identify putative candidates for the *Sk* gene and causative mutations for the bitter/sweet phenotype. Using the BPPCT017 (11Mb) and BPPCT038 (14.6Mb) markers flanking the *Sk* locus as the boundaries, reduced the 311,497 polymorphisms identified on scaffold 5 to 56,155 located between the SSR markers that have been reported previously as flanking the *Sk* locus (Table 4). Subsequent reduction of this dataset was completed by removing polymorphisms that were not homozygous in both sweet cultivars and within both bitter accessions. Also, the homozygous polymorphisms were required to be different between the sweet and bitter accessions yielding 6,304 polymorphisms of which 228 caused codon changing mutations. These missense, nonsense, and read-through SNPs; as well as the indels, comprise the reduced set of putative candidates for future screening and analysis.

Blast2GO Global Analysis

A global comparison of putative nonsense mutations within cherry and the four selected genotypes of almond reveal a similar distribution of mutations across various gene ontology terms. This can be seen in GO terms relating to biological process, molecular function as well as cellular component (Additional File 6). Response to stress, protein modification process, catabolic process, and transport each comprised at least 10% each of the total biological process GO terms for each tested dataset. With respect to molecular function, over 35% of annotated genes containing nonsense SNPs are involved in nucleotide binding with approximately 15% having kinase activity and slightly fewer than 15% having DNA binding activity. Finally, with respect to cellular component, about 25% of all annotated genes were predicted to be localized to the plastid, with both the mitochondrion and plasma membrane comprising 15% of all annotated genes. As there appeared to be little variation in the GO term composition of the five datasets, Blast2GO analysis of the entire peach gene set was performed and compared to datasets mapping back to nonsense-SNPs. A chi-square test

revealed that several GO-terms have statistically higher or lower GO terms than predicted (Additional File 7). Nonsense mutations map back to a total of 133 unique KEGG pathways, with Bitter 1 mapping to 121, Bitter 2 to 119, Sweet 1 to 124, Sweet 2 to 127, and Cherry to 127 KEGG pathways, respectively (Additional File 8). The cherry nonsense SNP-containing dataset contains members participating in atrazine degradation, chlorocyclohexane and chlorobenzene degradation, fluorobenzoate degradation, synthesis and degradation of ketone bodies, toluene degradation while none of the investigated almond accessions did. Conversely, all four almond genotypes contained predicted nonsense mutations within genes involved in butirosin and neomycin biosynthesis, D-alanine metabolism, D-arginine and D-ornithine metabolism. In the almond accessions, nonsense mutations were also found in genes assigned in databases as involved in glucosinolate biosynthesis all of which lacked participating genes with putative nonsense mutations in cherry. Since members of the Rosaceae family do not produce glucosinolates but synthesize cyanogenic glucosides using similar gene families the assignment of such genes to glucosinolate biosynthesis is obviously erroneous (Conn, 1969; Sánchez-Pérez et al., 2008).

Comparison of the nonsense-containing genes within the five data sets reveals that a large subset of the genes, 1,191 in total, is shared between all members (Figure 1). Additionally, some nonsense SNPs are unique to individual samples. The largest of these sets, 2,535 genes, are the nonsense mutations unique to cherry. 1,276 genes containing putative nonsense SNPs are present within each individual almond genotype and absent from the cherry analysis.

Blast2GO Targeted Pathway Analysis

The most abundant biological process gene ontology (GO) term represented in the datasets, “Response to stress” was selected as a GO of interest to further investigate. Further breakdown of this category reveals that its members are involved in a total of 92 KEGG pathways within the 5 investigated datasets (Additional File 9). While all datasets contain genes with putative nonsense SNPs in numerous pathways, only sweet cherry contains putative nonsense SNPs related to stress in C5-branched dibasic acid metabolism, chlorocyclohexane and chlorobenzene degradation, indole alkaloid biosynthesis, isoquinoline alkaloid biosynthesis, naphthalene biosynthesis, N-glycan biosynthesis, nicotinate and nicotinamide metabolism, primary bile acid biosynthesis, retinol metabolism, steroid degradation, steroid hormone biosynthesis, toluene degradation, tropane, piperidine and pyridine alkaloid biosynthesis and valine, leucine and isoleucine biosynthesis. Alternately, all four accessions of almond contain potential nonsense SNPs in alanine, aspartate and glutamate metabolism, benzoate degradation, caprolactam degradation, fatty acid elongation, geraniol degradation, monoterpenoid biosynthesis, sulfur metabolism, and vitamin B6 metabolism while cherry lacks nonsense mutation in these pathways.

Further investigation into nonsense mutations present within members of the genes involved in cyanogenic glucoside metabolism, was performed as these biosynthetic and catabolic pathways lead to amygdalin synthesis and catabolism, respectively. Blast2GO analysis performed through searching for the keywords “Prunasin” and “Amygdalin” revealed the presence of 4 isoforms of peach prunasin beta-glucosidase and amygdalin beta-glucosidase with nonsense mutations in cherry (ppa003891m, ppa016583, ppa003856m, and ppa003831m), one in Bitter 1 (ppa003831m), three in Bitter 2 (ppa003856m, ppa003891m, ppa003831m), four in Sweet 1 (ppa003891m, ppa016583, ppa003856m, and ppa003831m), and four in Sweet 2 (ppa003891m, ppa016583, ppa003856m, and ppa003831m). Based upon the annotations, numerous other members within this pathway contained putative nonsense mutations and additional members with potential prunasin beta-glucosidase or amygdalin beta-glucosidase activity were detected (Figure 2 and Table 5).

This data was further probed to identify differences in nonsense SNPs involved in ripening, another biological process of interest with respect to the *Prunus* genus. Within the 25 genes in all peach sequences mapping to the ripening GO term, cherry surprisingly contains 8 which are predicted to contain putative nonsense SNPs (two putative multidrug resistance genes, two 1-aminocyclopropane-1-carboxylate synthase (ACS) genes, two 1-aminocyclopropane-1-carboxylate oxidase (ACO) genes, one polygalacturonase gene and one glycosyl hydrolase family 9 protein gene) (Additional File 10). Bitter 1 and Bitter 2 almond datasets each contain two nonsense SNP-containing genes (one ACO gene and one putative multidrug resistance gene) while Sweet 1 and Sweet 2 almonds each contain three (two ACO gene and one putative multidrug resistance gene). Further analysis of genes containing SNPs in the KEGG pathway “cysteine and methionine biosynthesis” reveals additional nonsense SNPs that may affect ethylene synthesis, but are not mapped to the ripening GO term (Figure 3 and Table 6).

The final biological process of interest, abscission, has very few members containing nonsense mutations compared to the other GO terms, in fact only 23 genes in the entire peach gene set map to this term. Sweet 1 and Sweet 2 almonds are predicted to have a single gene related to abscission containing a nonsense SNP in a gene encoding btb/poz ankyrin repeat protein. Bitter 1 almond sequences had no nonsense mutations in any predicted gene sequences, while Bitter 2 almond has a nonsense mutation in a gene encoding a probable adp-ribosylation factor gtpase-activating protein (agd5-like). Cherry has nonsense SNPs present both in the gene encoding adp-ribosylation factor gtpase-activating protein (agd5-like) as well as the btb/poz ankyrin repeat protein.

Discussion

Uneven distribution of sequencing reads

Reference based assemblies are built upon the presumption that the sequenced genomes are highly similar to the reference. When differences exceed the threshold of the mapping software, reads from these highly divergent regions are not mapped. 7.5% of reads from *indica* rice cultivars did not map to the Nipponbare rice genome (Subbaiyan et al., 2012). In our data, 0.2% of the Illumina reads were unmapped to the peach genome (International_Peach_Genome_Initiative, 2013) strengthening the assertions that peach is a credible reference genome for *Prunus*.

In addition to unmapped reads, the data produced in this work identified 162 and 186 scaffolds of peach that were not covered by any reads from almond and sweet cherry respectively. One explanation for this is that these smaller, un-anchored scaffolds may be unique to the peach genome. Alternatively, as the peach genome was built using Sanger sequencing, these could be repetitive regions and the shorter reads used here were placed in the dual location during the mapping. In either case, these regions provide insights into genome structure differences that need to be further evaluated to fully understand the differences among these *Prunus* species.

Polymorphism Analyses

Our analyses show 48% of the mutations in the CDS to be non-synonymous, being either nonsense or missense mutations, in almond and 50% in sweet cherry. This is comparable to the 57% found in rice cultivars (Subbaiyan et al., 2012), however it is interesting that this interspecies comparison identified a higher percentage of synonymous SNPs than the intraspecific comparison in rice. While it is possible that the heterozygous nature of the almond and sweet cherry genotypes caused polymorphisms to be screened out during

filtering, it is unlikely that this would have significantly shifted the representation of synonymous and non-synonymous mutations.

It is important to note that the read-through mutations could be discussed as nonsense mutations of the almond gene in peach; therefore, discussion of read-through and nonsense mutations is limited by the perspective of the analysis which, in this case, is in respect to the peach reference genome. At first glance, the 0.1% generation rate of read-through mutations suggests that these mutations may be highly deleterious with strong selection against them as they occur at $\sim 1/50^{\text{th}}$ of the rate that nonsense mutations arise. A closer examination, however, reveals that while the probability of a stop codon mutation causing a read-through mutation is 85%, there is only one stop codon per protein. This contrasts significantly with the 4.2% chance of a random SNP causing a nonsense mutation multiplied by the 403 amino acids found in the average CDS in the peach genome. Calculating for the distribution of amino acids yields one polymorphism having a 4.18% or 0.21% chance of causing a nonsense or read-through mutation respectively in the average peach gene. The data from this work show a 2.5 fold change from expected providing intrigue but requiring further evaluation regarding the effect of these mutations on gene function.

346 regions with higher and lower rates of polymorphism were identified in this work. Higher rates could result from genomic duplications or from low conservation yielding more divergence. Similarly, regions with lower than average polymorphisms could be the result of either low divergence where few polymorphisms arose, or of very high amounts of differentiation preventing the mapping of the sequencing reads to these locations. Subbaiyan (2012) revealed similar regions of lower polymorphism rates in six inbred lines of rice with several being greater than 100kb in length. The 600kb region in almond is particularly interesting as it may represent a larger region of diversity between almond and peach and may contain genes related to the divergence of these two species.

Analysis of the *Sk* locus

The combination of the existing DNA markers, the reference sequence, and genotype specific sequencing yielded 228 candidate mutations for the sweet kernel trait in almond. Since this work was completed using only 2 bitter and 2 sweet genotypes, reductions in this candidate set would be expected if more genotypes were examined. However, whole genome sequencing of further genotypes is not necessary at this time as site specific testing of the genotypes for the identified mutations are expected to identify the allele responsible for the difference between these types of almonds. As the major and highly critical trait, developing a gene based marker for the sweet kernel gene will provide an important benefit to the almond community by rapidly identifying the undesirable bitter genotypes. As suggested by Micheltore (1991), bulked segregant analysis can function in an obligate outcrossing species. The results shown here demonstrate the ability of the approach to produce a small candidate list from a large region of interest. Adding more individuals to the bulks in this work would allow the marker placement to be independently confirmed as well, though using 2 individuals of each phenotype was possible due to the previously developed markers for the sweet kernel locus.

Blast2GO comparisons

The global distribution of GO terms within the nonsense SNP-containing genes was similar among all samples tested. This suggests two potential options regarding the presence of nonsense SNPs: 1) Certain gene ontology terms have accrued nonsense SNPs at similar levels across species in *Prunus* or 2) Nonsense SNPs simply occur randomly throughout the genome. Each gene ontology term contains a similar number of genes among the samples investigated. In order to assess this, comparison of the observed number of members for

each GO term was performed against expected values generated from the entire peach predicted gene set using a chi-square test. This test showed that numerous GO categories contained statistically significant higher or lower numbers of nonsense SNPs than expected (Additional File 7). This suggests that many GO terms are linked to an increased likelihood to generate nonsense SNPs in *Prunus*, while other GO terms appear to be more conserved in the genus supporting option 1 above. Interestingly, the GO terms associated with significantly higher nonsense SNPs ($p\text{-value} < 1E^{-10}$) include the biological processes “DNA metabolic process” (GO:0006259), “cellular protein modification process” (GO:0006464), “signal transduction” (GO:0007165), and “pollen-pistil interaction” (GO:0009875); the cellular components “mitochondrion” (GO:0005739), “cytoskeleton” (GO:0005856), and “plastid” (GO:0009536); and the molecular functions “nucleotide binding” (GO:0000166) and “kinase activity” (GO:0016301). GO terms associated with significantly lower nonsense SNPs ($p\text{-value} < 1E^{-10}$) include the biological processes “response to biotic stimulus” (GO:0009607), “response to abiotic stimulus” (GO:0009628), “anatomical structure morphogenesis” (GO:0009653) and “response to endogenous stimulus” (GO:0009719); the cellular component “cytosol” (GO:0005829); and the molecular functions “chromatin binding” (GO:0003682), “sequence-specific DNA binding transcription factor activity” (GO:0003700) and “structural molecule activity” (GO:0005198). While a connection between GO term and occurrence of nonsense SNPs appears to exist, this does not disprove the option 2 stated above.

Concerning the GO term “response to stress”, there appears to be significant genetic variability with respect to nonsense SNPs. In fact, sequences containing detected nonsense SNPs mapped to this GO term more than any other GO term investigated in the biological process domain. This gene ontology is of high agricultural importance as breeding and genetic modification of plants resistant to both biotic and abiotic stresses is a large focus in both industry and academia. Previous studies have used gene-based SNPs detected through interspecific comparisons to identify, verify and attach function to SNPs which may be involved in stress response (Parida et al., 2012). These putative nonsense SNPs represent a preliminary dataset within *Prunus* which may be used in similar studies.

Basic differences exist in the ripening patterns of members of the *Prunus* genus. Peach, apricot, and plum fruits are climacteric, meaning that a burst of ethylene occurs quickly followed by an increase in respiration. Cherry and almond, on the other hand exhibit non-climacteric ripening, an outlier in the genus. The identification of nonsense mutations in several versions of ACS and ACO could significantly disrupt the ethylene production pathway in cherry rendering it nearly unable to provide the burst seen in other fruits in this genus. While cherry is non-climacteric, the cherry fruit color and maturation is modulated by application of exogenous ethylene (Koepke and Dhingra unpublished).

While these results enable the identification of targets for gene linked marker screening, it is important to realize the limitations of this project. First of all, nonsense SNPs do not necessarily equate to loss of function of a protein. Additionally, as these sequences were aligned to a predicted peach data set, the true sequence of genes of interest may be biased. Potential splice variants may have the ‘nonsense’ mutation in an exon that is not utilized in these species. Also, the presence of a single nonsense mutation may not be deleterious at all and could be sufficiently complemented by the other allele especially in a genus where very few self-compatible varieties exist leading to high amounts of heterosis. Gene duplications or those genes unique to almond or cherry may not be represented in these data; alternatively, they may be represented as SNPs while they are actually different alleles.

Conclusions

Using reference based assemblies of four almond accessions and one sweet cherry cultivar, we were able to begin interspecific comparative genomic analysis of Prunoideae. Over 99% of the raw reads mapped to the peach genome though nearly 44% mapped to the chloroplast. Identifying hundreds of smaller scaffolds in the peach genome that were not mapped to by either the almond or sweet cherry data identifies many potentially peach-specific regions of interest for further investigation. The 6.1 million putative SNPs provide a resource for gene based investigations. While many of the SNPs and indels are in non-coding regions, 250 to 300 thousand SNPs are located in the coding regions of annotated peach genes. These SNPs should prove to be useful in expanding our knowledge of genetics and genomics in these species through their use as molecular markers and gene based interrogations. The coverage depth images revealed 31 regions that have significantly different amounts of SNPs.

A keystone goal of genomics is to identify genes responsible for specific traits. Here, we examined the bitterness trait of almond and identified 228 codon-changing mutations near the previously identified *Sk* locus. Additionally, to the best of our knowledge, we provide the first report in plants of nonsense SNP abundance in a genus being linked to specific GO terms. A global analysis of SNPs has also revealed several candidate mutations of interest for different physiological properties of these species including response to stress, ripening and abscission. Combined, these data should provide a foundation for further genomics and genetics research in Prunoideae.

Methods

Sequencing data acquisition

Almond—D05-187 (Bitter1) and S3067 (Bitter2) are homozygous bitter selections from the CEBAS-CSIC and Ramillete (Sweet1) and Lauranne (Sweet2) are each homozygous sweet cultivars of almond. Using an estimated genome size of 250Mb, approximately 10x coverage was obtained for each of the four genotypes with 76bp Illumina paired-end reads.

Cherry—The sweet cherry genome project has developed roughly 7x coverage of Stella, an important parental cultivar based on a 225Mb genome size estimation. These data were derived mostly through single-end 454 with some paired-end 454 and Illumina paired-end sequencing. Both 454 GS-FLX and 454 GS-FLX+ versions were used to acquire these sequences. Also, 454 transcriptome data from Bing and Rainier cultivars of sweet cherry were obtained and used in the analyses. These transcriptome data were utilized only for polymorphism analysis to compare to the gene annotations of peach and were not obtained in sufficient depth for expression based analyses.

Peach Genome—The peach genome version 1.0 (International_Peach_Genome_Initiative, 2013) was obtained from GDR (Jung et al., 2008) for use as the reference sequence throughout this project. The chloroplast and mitochondrial genomes were excluded from the assembly initially and the chloroplast was later used to screen the unassembled reads.

Assembly

A reference based assemblies of both the cherry genomic 454 and cherry transcriptomic 454 data was assembled using the NGen assembler (DNASTar) version 3.1.0 with the peach genome version 1.0 as the reference and using the following 454 default parameters: mersize =21, merSkipQuery = 3, minMatchPercent = 85, MaxGap = 15, minAlignedLength = 50. Similarly, all Illumina data from the four almond accessions and sweet cherry were

assembled using the peach genome as a reference with the Illumina default parameters: mersize = 21, minMatchPercent = 93, mismatchPenalty = 20, MaxGap = 6, min Aligned Length = 35. For each assembly, the different genotypes were input separately to enable unique SNP information to be attained for each individual.

Polymorphism Analyses

Assembled data were imported into SeqMan (DNASTar) where SNP reports were created. A custom script was used to remove polymorphisms with less than 3 reads confirming each non-reference call similar to previous SNP reporting works (Deschamps and Campbell, 2010; Hyten et al., 2010; Koepke et al., 2012; Kulheim et al., 2009). These filtered SNPs were then imported into ArrayStar (DNASTar) to enable further analyses.

Polymorphism type and region identification—Custom computational comparisons of the base calls from the sequenced individuals against the peach genome were completed to determine the base changes involved. Similarly, polymorphism regions were identified by analyzing the reference position against the annotation of the peach genome. These SNPs were classified as 5' UTR, intron, exon, 3' UTR or intergenic. Exonic polymorphisms were further classified as sense, nonsense, missense or read-through mutations based on the resulting amino acid compared to the peach genome annotation. Read-through mutations were defined as the SNPs causing a stop codon to be changed into an amino acid thereby elongating the C terminus of the protein with respect to the peach gene (Zirn et al., 2005).

Polymorphism depth analyses—To visualize the depth of the polymorphisms across the 8 main scaffolds of the peach reference, the total polymorphisms in each discrete 50kb window were analyzed and displayed as a single pixel wide bar one pixel high for each 20 polymorphisms. The graphs for each individual were then compiled into a single image per scaffold. The composite polymorphism set, where each unique SNP was counted once for each species, was also analyzed in this manner. The distribution of polymorphism counts per 50kb window was analyzed to identify regions of the peach reference that had a polymorphism depth greater than 2 standard deviations from the mean of that scaffold.

Analysis of *Sk* locus

The total almond SNP report was filtered to retain only the polymorphic sites near the *Sk* locus. Since the markers BPPCT017 and BPPCT038 are located at ~11Mb and 14.6Mb on peach linkage group 5 respectively (Sánchez-Pérez et al., 2010), they were used for the bounds around the *Sk* locus. All polymorphisms that were conserved within a group but contrasting between the two types were retained as both bitter and both sweet genotypes are homozygous for the trait. Further screening reduced the data set to only contain codon changing polymorphisms that make up the candidate gene set.

Blast2GO comparisons

Nucleotide sequences for all predicted *Prunus persica* genes were imported into Blast2GO (Conesa et al., 2005; Gotz et al., 2008). Details of Blast2GO methods used are provided in Additional File 11 z. A Gene Annotation File containing the information from this study was submitted to the Plant Ontology project. A chi-square test was performed to determine if the observed GO distribution of nonsense SNP-containing genes was significantly different from the expected. Custom scripts were used to compare datasets to determine which contained unique or shared entries. Finally, KEGG pathway maps and corresponding information were downloaded from the KEGG Pathway Database through Blast2GO (<http://www.genome.jp/kegg/pathway.html>) (Kanehisa, 2002; Kanehisa et al., 2012).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

AD and NO would like to acknowledge the support received from WSU ARC startup and Hatch funds for this project. Washington Tree Fruit Research Commission support to AD and NO is gratefully acknowledged. TK and SS acknowledge support received from NIH Protein Biotechnology Training Program T32GM008336 and ARCS fellowships. AH was supported in part by a US Department of Agriculture National Research Initiative (USDA-NRI) grant 2008-35300-04676 to AD. RSP would like to thank “Séneca Foundation” for the project “Molecular Biology of Cyanogenesis in Almonds” and “MINECO” for the project “Mejora Genética del Almendro”. RSP is also grateful for her postdoctoral contracts by CSIC (JAE Doc) and MINECO. HS is supported by CONICYT, FONDECYT/Regular N°1120261 and Innova CORFO (07CN13 PBT-167). LM is supported by CONICYT, FONDECYT/Regular N°1121021 and Innova CORFO (07CN13 PBT-167). BLM would like to thank the Villum research center Pro-Active Plants and the Novo Nordisk Foundation Center for Bio-Sustainability for financial support.

Works Cited

- Ahmad R, Parfitt DE, Fass J, Ogundiwin E, Dhingra A, Gradziel TM, Lin DW, Joshi NA, Martinez-Garcia PJ, Crisosto CH. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *Bmc Genomics*. 2011; 12
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21:3674–3676. [PubMed: 16081474]
- Conn EE. Cyanogenic glycosides. *J Agr Food Chem*. 1969; 17:519–526.
- Deschamps S, Campbell M. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Molecular Breeding*. 2010; 25:553–570.
- Dicenta F, García JE. Inheritance of the Kernel Flavor in Almond. *Heredity*. 1993; 70:308–312.
- Dicenta F, Ortega E, Martínez-Gómez P. Use of recessive homozygous genotypes to assess genetic control of kernel bitterness in almond. *Euphytica*. 2007; 153:221–225.
- Ganal MW, Altmann T, Roder MS. SNP identification in crop plants. *Current Opinion in Plant Biology*. 2009; 12:211–217. [PubMed: 19186095]
- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008; 36:3420–3435. [PubMed: 18445632]
- Hyten DL, Cannon SB, Song QJ, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *Bmc Genomics*. 2010; 11
- International_Peach_Genome_Initiative. The high quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. 2013
- Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D. GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. *Nucleic Acids Res*. 2008; 36:D1034–D1040. [PubMed: 17932055]
- Kanehisa M. The KEGG database. In *Silico Simulation of Biological Processes*. 2002; 247:91–103.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012; 40:D109–D114. [PubMed: 22080510]
- Koepke T, Schaeffer S, Krishnan V, Jiwan D, Harper A, Whiting M, Oraguzie N, Dhingra A. Rapid gene-based SNP and haplotype marker development in non-model eukaryotes using 3' UTR sequencing. *Bmc Genomics*. 2012; 13
- Kulheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF. Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *Bmc Genomics*. 2009; 10

- Michelmore RW, Paran I, Kesseli RV. Identification of Markers Linked to Disease-Resistance Genes by Bulk Segregant Analysis - a Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. *Proceedings of the National Academy of Sciences of the United States of America*. 1991; 88:9828–9832. [PubMed: 1682921]
- NASS U. Noncitrus Fruits and Nuts 2010 Summary. 2011
- Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007
- Parida SK, Mukerji M, Singh AK, Singh NK, Mohapatra T. SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. *Bmc Genomics*. 2012; 13:426. [PubMed: 22921105]
- Sánchez-Pérez R, Belmonte FS, Borch J, Dicenta F, Moller BL, Jorgensen K. Prunasin Hydrolases during Fruit Development in Sweet and Bitter Almonds. *Plant Physiology*. 2012; 158:1916–1932. [PubMed: 22353576]
- Sánchez-Pérez R, Howad W, Garcia-Mas J, Arus P, Martinez-Gomez P, Dicenta F. Molecular markers for kernel bitterness in almond. *Tree Genetics & Genomes*. 2010; 6:237–245.
- Sánchez-Pérez R, Jørgensen K, Olsen CE, Dicenta F, Moller BL. Bitterness in almonds. *Plant Physiology*. 2008; 146:1040–1052. [PubMed: 18192442]
- Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arus P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE. Multiple models for Rosaceae genomics. *Plant Physiology*. 2008; 147:985–1003. [PubMed: 18487361]
- Subbaiyan GK, Waters DLE, Katiyar SK, Sadananda AR, Vaddadi S, Henry RJ. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal*. 2012:no–no.
- Zhebentyayeva TN, Swire-Clark G, Georgi LL, Garay L, Jung S, Forrest S, Blenda AV, Blackmon B, Mook J, Horn R, Howad W, Arus P, Main D, Tomkins JP, Sosinski B, Baird WV, Reighard GL, Abbott AG. A framework physical map for peach, a model Rosaceae species. *Tree Genetics & Genomes*. 2008; 4:745–756.
- Zirn B, Wittmann S, Gessler M. Novel Familial WT1 Read-Through Mutation Associated With Wilms Tumor and Slow Progressive Nephropathy. *American Journal of Kidney Diseases*. 2005; 45:1100–1104. [PubMed: 15957141]

Figure 1.

Venn diagram displaying presence of nonsense SNPs present within the five investigated datasets mapped against peach predicted genes. A comparison of the composition of putative nonsense SNP containing genes between the four investigated genotypes of almond and the combined cherry data set reveals the presence of a large set, 1,191, of nonsense containing homologues across all members. Additionally, each sample has a unique set of genes containing putative nonsense SNPs, most notably cherry with 2,535 genes.

Figure 2. KEGG Pathway of genes containing nonsense mutations within cherry and almond samples involved in Cyanogenic glucoside metabolism. Table 5 lists the EC numbers, their identities, as well as the predicted CDS from *Prunus persica* which contain a SNP for each respective species. Colored boxes correspond to the color code found in the first column of Table 5.

Figure 3. KEGG pathway of genes containing nonsense mutations within cherry and almond samples participating in cysteine and methionine metabolism. Genes containing nonsense mutations were mapped to EC numbers using Blast2GO and mapped on the cysteine and methionine metabolism KEGG map closest to ethylene production. Table 6 lists the identity of each enzyme, as well as the predicted CDS from *Prunus persica* which contain a SNP for each respective species. Only the lower half of the pathway was imaged. Colored boxes correspond to the color code found in the first column of Table 6.

Raw sequencing data

Table 1

Total data acquired for one sweet cherry cultivar and four almond accessions. Only genomic data was used for almond genotypes

	Sweet cherry		Almond			
	Transcriptome	Genomic	Bitter1	Bitter2	Sweet1	Sweet2
Data type	454	454 Illumina	Illumina	Illumina	Illumina	Illumina
Total Sequences:	1,225,030	3,742,780 977,713	29,202,304	43,522,066	42,403,474	27,607,822
Total Bases (Mb):	467	1,020 557	2,219	3,307	3,222	2,098
Mean read length:	381	272 57	76	76	76	76
Approximate genome coverage	2.1x	4.5x 2.5x	8.9x	13.2x	12.9x	8.4x

Table 2

Nonsense mutation analysis statistics.

	Sweet Cherry	Bitter 1 Almond	Bitter 2 Almond	Sweet 1 Almond	Sweet 2 Almond
Total Number of Peach Genes Represented	27,576 (96.20%)	28,332 (99.33%)	28,420 (99.64%)	28,488 (99.38%)	27,590 (96.73%)
Total Genes with Predicted Nonsense Mutation(s)	5,384	4,016	5,110	5,302	5,529
Total Genes with Predicted Nonsense Mutation(s) Unique to Each Genotype	2,535	190	409	467	624

Table 3

Classifications of polymorphisms identified in each data set based on their location relative to peach genes. Locations are denoted based on the annotations as found in Additional File 2. Potential synonyms for the mutations are listed in parentheses.

Mutation Location (Synonym)	Cherry			Almond			
	Cherry Transcripts	Cherry 454	Cherry Illumina	Almond Sweet1	Almond Sweet2	Almond Bitter1	Almond Bitter2
Gene Total	247,818	701,534	39,338	678,587	774,780	483,905	646,641
mRNA Total	266,340	739,110	40,978	707,564	808,159	504,028	673,423
CDS - Total	194,279	305,314	22,881	274,904	290,793	207,911	266,252
CDS - Sense	99,947	12,702	149,960	132,747	156,693	196,849	133,753
CDS - Mis-sense	80,131	10,688	127,288	118,544	151,659	106,056	138,853
CDS - Nonsense (Stop Gain)	3,261	443	20,791	32,810	8,109	5,446	7,452
CDS - Read-through (Stop loss)	113	23	312	355	404	257	354
CDS - Deletions	5,114	6,415	107	2,415	3,399	3,549	4,028
CDS - Insertions	4,820	9,533	100	2,299	3,295	3,412	3,761
3' UTR	32,317	42,486	1,083	26,924	30,330	19,015	25,398
5' UTR	10,788	13,967	402	13,309	14,687	10,102	12,891
Intergenic	43,137	797,699	58,654	1,789,323	2,263,108	1,208,734	1,734,295

Table 4

Sk locus analysis demonstrating the effect of the various parameters on the reduction in potential targets related to bitterness in almond.

	Number of target polymorphisms
A. Chromosome 5	311,497
B. A + 11–14.6MB	56,155
C. B + fitting genetic patterns	6,304
D. C + with codon change	228

Table 5
Prunus persica predicted CDSIDs containing nonsense mutations for putative members in cyanoamino acid metabolism (Figure 2).

EC Number	Enzyme	Cherry	Bitter 1	Bitter 2	Sweet 1	Sweet 2
3.2.1.21 (Red)	Beta-glucosidase	ppa018777m ppa015330m ppa004484m ppa001692m ppa006142m ppb011574m ppa023763m ppa001675m	ppa015619m ppa001675m ppa004484m ppa006167m ppa006142m ppa023264m ppa023763m	ppa015619m ppa001675m ppa004484m ppa001656m ppa006167m ppb021184m ppa026252m ppa024207m ppa020836m ppa023264m ppa023763m	ppa015619m ppa014607m ppa018777m ppa019582m ppa001675m ppa001656m ppa006167m ppa004484m ppa001656m ppa006167m ppa007195m ppb021184m ppa020836m ppa024207m ppa023264m ppa020836m ppa023264m ppa023763m	ppa015619m ppa015239m ppa014607m ppa014605m ppa018777m ppa019582m ppa001675m ppa004484m ppa001656m ppa006167m ppa007195m ppb021184m ppa020836m ppa024207m ppa023264m ppa020836m ppa023264m ppa023763m
3.5.5.4 (Yellow)	Cyanoalanine nitrilase	ppa008090m	ppa008090m	ppa008090m	ppa008090m	ppa008090m
3.5.5.1 (Green)	Nitrilase	ppa008583m	ppa008767m	ppa008767m	ppa008767m	ppa008767m ppa007102m
6.3.1.1 (Orange)	Aspartate-ammonia ligase	0	ppa015268m	0	0	ppa015268m
4.1.2.10 (Brown)	(R)-mandelonitrile lyase	ppa016983m ppa003595m ppa003414m ppa003422m ppa004308m ppa020579m	ppa003595m ppa003414m ppa003422m ppa020579m	ppa003595m ppa003414m ppa003422m ppa020579m ppa022916m	ppa003595m ppa003414m ppa003422m ppa020579m ppa022916m	ppa003414m ppa003422m ppa020579m ppa022916m
3.2.1.118 (Blue)	Prunasin beta-glucosidase	ppa017484m ppa019137m ppa019262m ppa015721m ppa003718m ppa003891m ppa021158m ppa020817m ppa026358m ppa016583m ppa003856m ppa003831m	ppa017484m ppa019137m ppa018933m ppa018404m ppa003831m	ppa017484m ppa018933m ppa019262m ppa015161m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020817m ppa020368m ppa026358m ppa027189m	ppa017484m ppa016583m ppa015970m ppa019262m ppa016757m ppa015161m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020817m ppa020368m ppa026358m ppa027189m	ppa017484m ppa019137m ppa018933m ppa016583m ppa015970m ppa016757m ppa015161m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020817m ppa020368m ppa026358m ppa027189m

EC Number	Enzyme	Cherry	Bitter 1	Bitter 2	Sweet 1	Sweet 2
3.2.1.117 (Pink)	Amygdalin beta-glucosidase	ppa017484m ppa019573m ppa019137m ppa018933m ppa019262m ppa015721m ppa003718m ppa003891m ppa021158m ppa020817m ppa026358m ppa016583m ppa003856m ppa003831m	ppa017484m ppa019573m ppa019137m ppa018933m ppa019262m ppa018404m ppa004380m ppa003831m	ppa017484m ppa019573m ppa018933m ppa015970m ppa019262m ppa015161m ppa004380m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020817m ppa020067m ppa020368m ppa026358m ppa027189m	ppa017484m ppa019573m ppa016583m ppa015970m ppa019262m ppa016757m ppa015161m ppa004380m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020067m ppa025067m ppa020368m ppa026358m ppa027189m	ppa017484m ppa019573m ppa019137m ppa018933m ppa016583m ppa015970m ppa016757m ppa015161m ppa004380m ppa003718m ppa003856m ppa003891m ppa003831m ppa022831m ppa020067m ppa025067m ppa020368m ppa026358m ppa027189m
3.5.1.1 (Grey)	Asparaginase	ppa008583m	0	ppa016260m ppa008761m	ppa008583m	ppa016260m ppa008583m
1.14.13.68 (Purple)	4-hydroxyphenylacetaldehyde oximemonooxygenase	0	0	0	ppa014661m	0
2.1.2.1 (Cyan)	glycine hydroxymethyltransferase	ppa003640m	0	0	0	ppa004090m

Table 6

Prunus persica predicted CDSIDs containing nonsense mutations with putative functions cysteine and methionine metabolism (Figure 3).

EC Number	Enzyme	Cherry	Bitter 1	Bitter 2	Sweet 1	Sweet 2
2.6.1.57 (Red)	Aromatic-amino-acid transaminase	ppa003908m	ppa004475m	ppa004475m	ppa004475m	ppa004475m
4.4.1.14 (Yellow)	l-aminocyclopropane-1-carboxylate synthase	ppa015636m ppa016458m ppa004774m ppa003908m ppa005521m	ppa015636m ppa004475m	ppa015636m ppa004475m ppa003850m	ppa015636m ppa004475m ppa005521m	ppa015636m ppa004475m ppa003850m
1.14.17.4 (Green)	Aminocyclopropanecarboxylate oxidase	ppa008813m ppa008791m	ppa008813m	ppa008813m	ppa008813m ppa008791m	ppa008813m ppa008791m
6.3.1.1 (Orange)	Methionine synthase	0	ppa015268m	0	0	ppa015268m
4.1.1.50 (Lime Green)	Adenosylmethionine decarboxylase	ppa007732m	ppa007732m	0	0	ppa007294m
2.6.1.5 (Blue)	Tyrosine transaminase	ppa019805m	0	ppa019805m	ppa019805m	ppa019805m ppa018754m
2.1.1.10 (Pink)	Homocysteine S-methyltransferase	ppa008404m	0	0	0	ppa010310m
2.1.1.37 (Grey)	DNA (cytosine-5-)methyltransferase	ppa019831m ppa000190m ppa006086m	ppa019831m	ppa015623m ppa000190m ppa006086m	ppa019831m ppa015623m ppa000190m ppa006086m	ppa000190m ppa006086m
2.5.1.6 (Purple)	Methionine adenosyltransferase	ppa006915m	0	0	ppa025497m	0
2.1.1.14 (Cyan)	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	0	ppa026306m ppa021650m	ppa021650m	ppa026306m ppa021650m	ppa021650m