

Optical mapping and sequencing of ethanologenic *E. coli* strain KO11 genome reveals 20-copy tandem duplication of *Z. mobilis pdc* and *adhB* genes

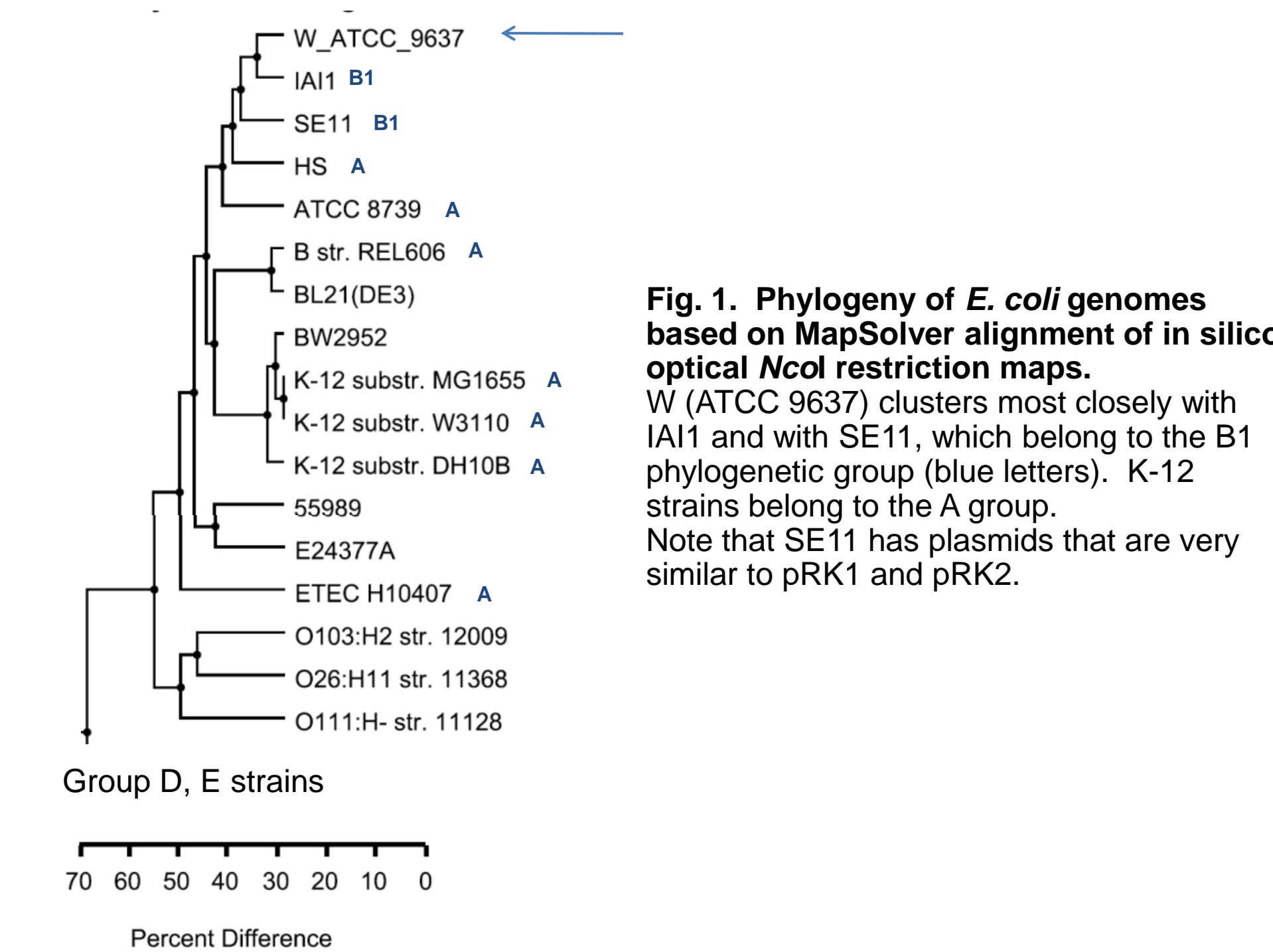
Lonnie O. Ingram¹, Peter C. Turner¹, Lorraine Yomano¹, Laura Jarboe¹, Christy Baggett¹, Brelan Moritz¹, Emily Zentz² and K. T. Shanmugam¹

¹Department of Microbiology & Cell Science, University of Florida, Gainesville, FL ²OpGen, Inc., Gaithersburg, MD

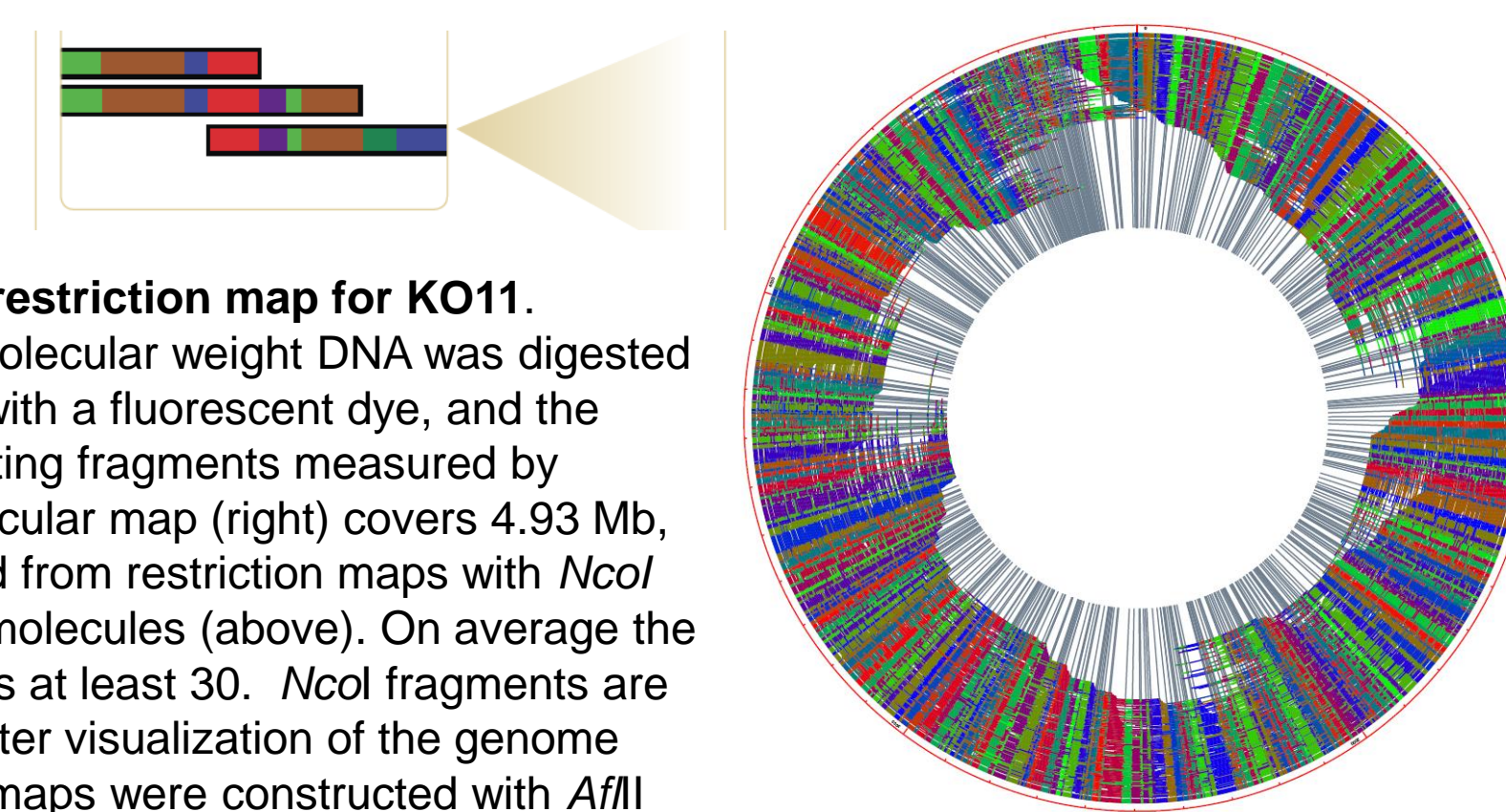
ABSTRACT

E. coli KO11 was engineered to produce ethanol by inserting the pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*) genes from *Zymomonas mobilis*. Production of succinate, an unwanted byproduct, was eliminated by mutating fumarate reductase (*frd*). The KO11 genome was sequenced by 454 technology and assembled with Newbler to produce 196 contigs. An optical *NcoI* restriction map of the entire KO11 genome was generated to enable closure of the contigs. Gaps were filled by sequencing PCR products extending between adjoining contig ends. The optical map of KO11 showed an apparent tandem repeat region, which was confirmed by additional *AflI* and *BamHI* maps of KO11. Each repeat was ~10 kb in length, and present in approximately 20 copies. The repeat region was in between contigs that flanked the *pflB* gene, which was used as the insertion site for the *pdc*, *adhB* and chloramphenicol resistance (*cat*) genes. The sequence coverage of contigs lying within the *pdc*, *adhB* and *cat* genes was about 20-fold higher than the average sequence coverage, consistent with tandem duplication of the foreign genes, which were inserted as a circularized DNA fragment. Selection for higher levels of chloramphenicol-resistance originally allowed isolation of strains with higher expression of both *pdc* and *adhB*, and hence improved fermentation performance, by increasing the number of gene copies. The chromosomal sequence of the parent strain ATCC 9637 (W) has also been completely assembled, showing extensive rearrangements in KO11, and enabling detection of changes in KO11 that may contribute to improved fermentation properties.

A comparison of the *NcoI* optical map of W with the predicted *NcoI* maps of all available complete *E. coli* genomes was made with MapSolver software (OpGen), yielding the tree shown below.



The KO11 genome is 5,021,750 bases in length, and carries plasmid pRK2 but not pRK1. Assembling KO11 contigs was difficult, but optical restriction maps of the KO11 chromosome allowed completion of the assembly process. The *NcoI* map for KO11 is shown below.



A comparison of the *NcoI* optical maps for W and KO11 (Fig. 3) revealed an unexpectedly large insertion in KO11 at the site where foreign genes had been introduced. Optical mapping of KO11 with *AflI* and *BamHI* (Fig. 3) showed that the insertion consisted of approximately 20 copies of a 10 kb repeat.

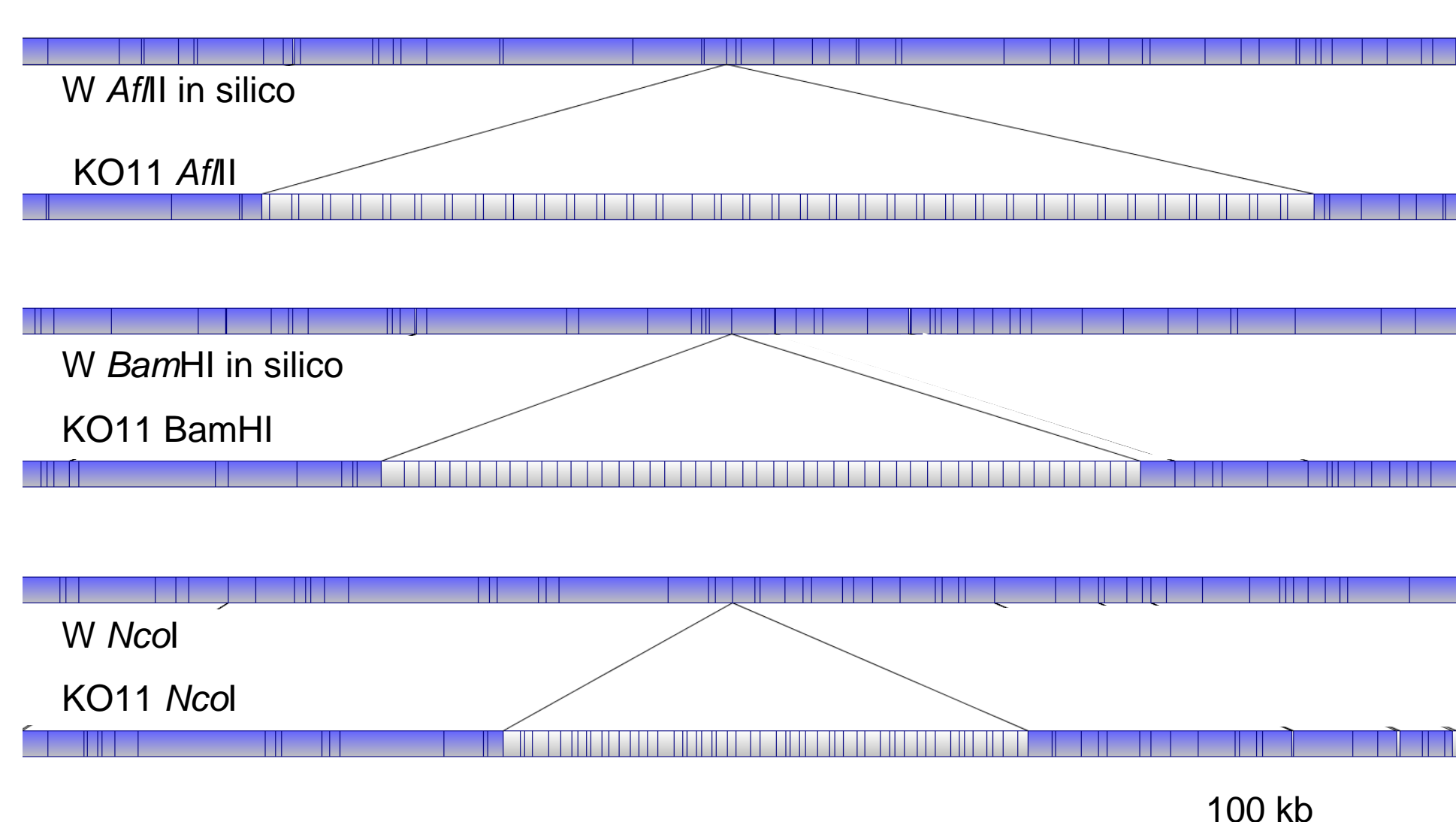


Figure 3. Optical maps for W (ATCC 9637) and KO11 aligned using MapSolver (OpGen), showing the insertion of the tandem repeat region (~20 copies) in KO11.
The top 2 lines are *AflI* maps; the middle two *BamHI*; and the lower two *NcoI*. Note the regular repeat structure of the inserted region evident from the *AflI* and *BamHI* maps.

Sequence coverage for contigs including the *pdc*, *adhB*, and *cat* genes was higher than average sequence coverage by a factor of ~25, consistent with a tandem repeat of these genes (Fig. 4). Amplification of the *cat* gene in response to selection for higher levels of Cam-resistance presumably resulted in concomitant amplification and increased expression of the *pdc* and *adhB* genes.

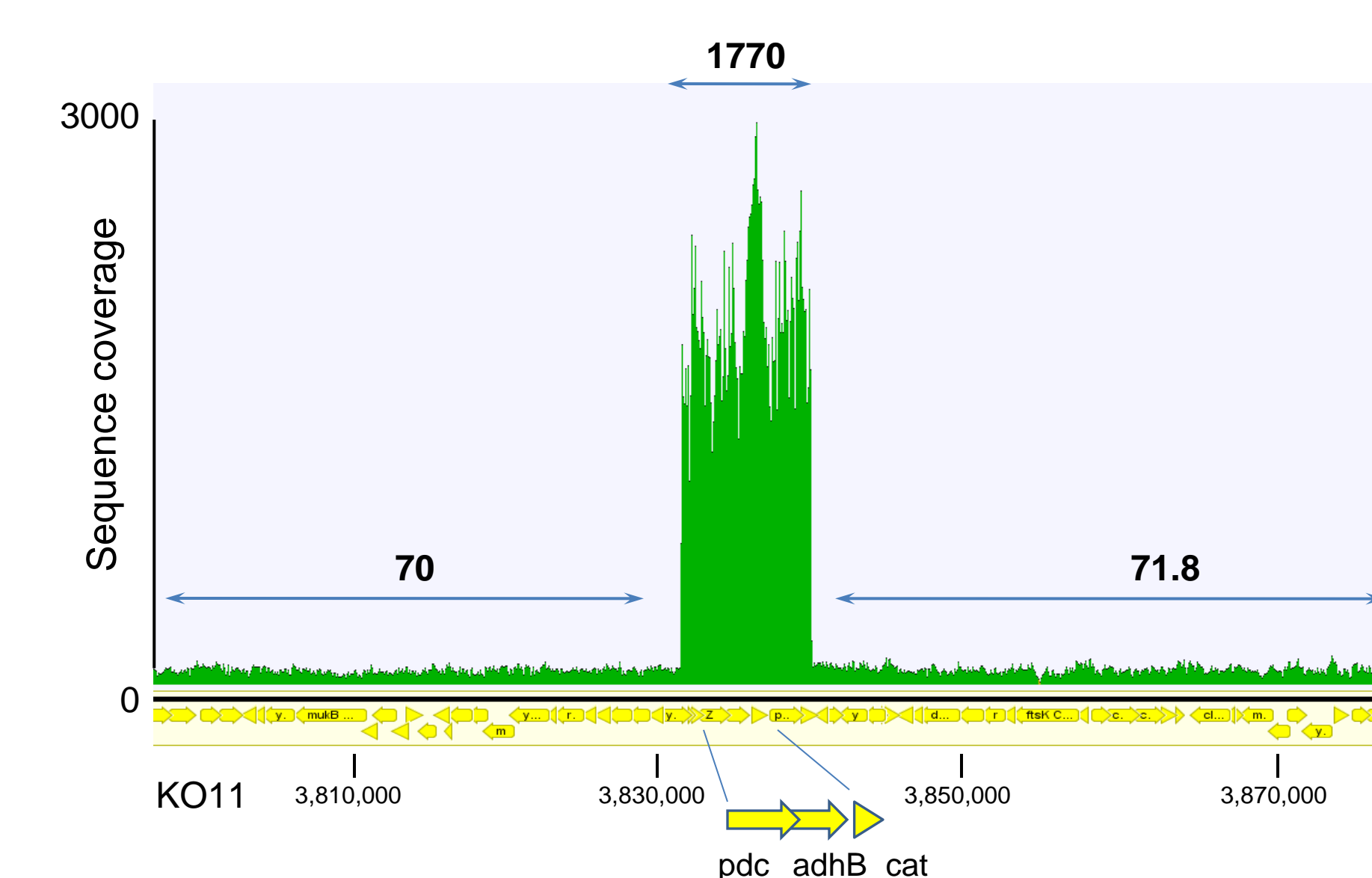


Figure 4. Sequence coverage over the KO11 region including the *Z. mobilis pdc* and *adhB* genes.
A template containing only one copy of the foreign inserted genes was used for this assembly. Numbers at the bottom are coordinates on the KO11 genome. The numbers over the blue arrows show the average sequence coverage for the indicated regions. Note the high sequence coverage (1770) for the *pdc*-*adhB*-*cat* region, approximately 25-fold higher than coverage to the left and right.

KO11 contains multiple copies of the 1.3 kb insertion sequence IS10, but *E. coli W* has no IS10 insertions. Transposon Tn10 containing IS10 at each end was introduced with K-12 DNA into KO11 during strain construction. The initial KO11 sequenced by the DoE JGI [5] contains only one IS10. However, our KO11 sequence reveals 30 different sites for IS10 (Fig. 5).

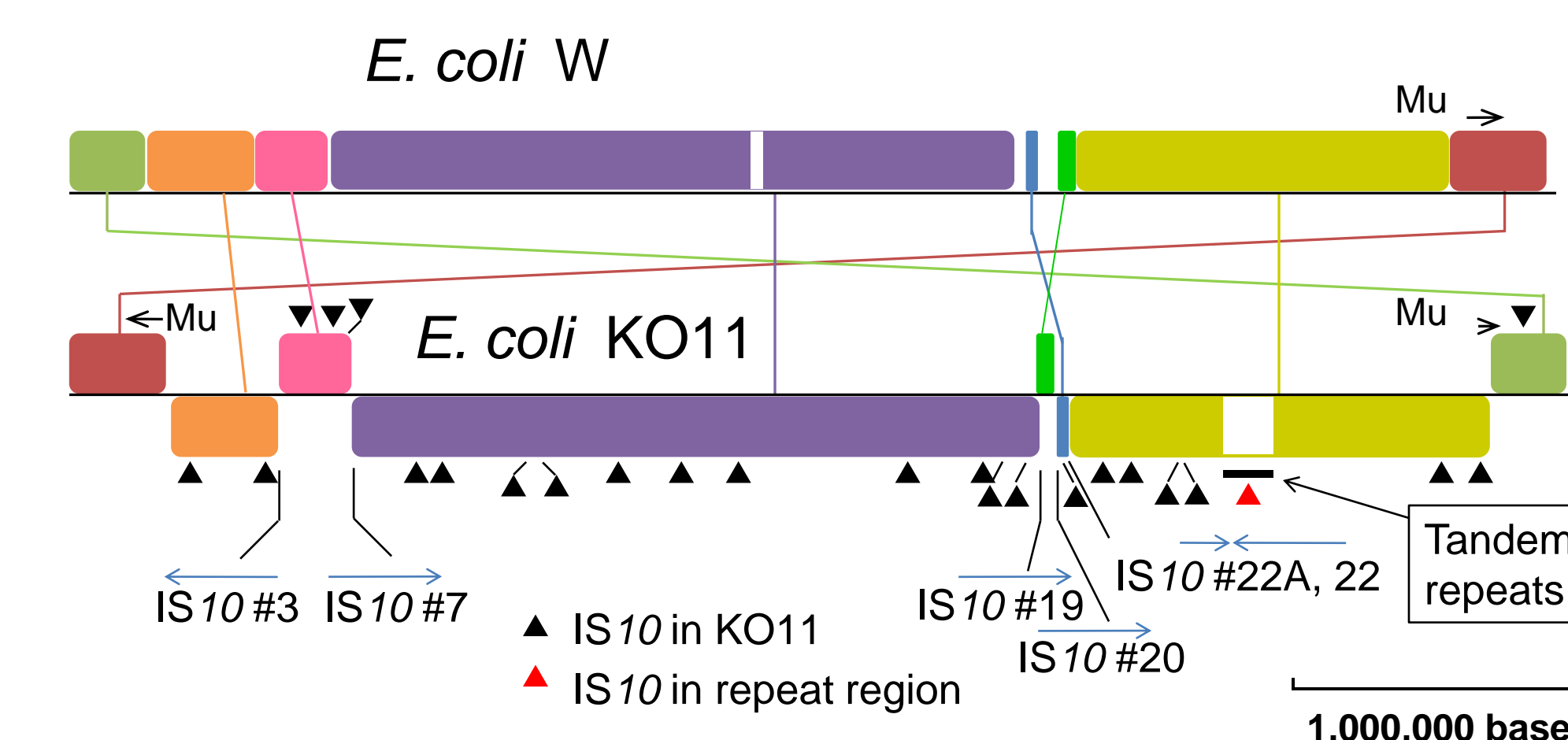


Figure 5. Alignment of W (ATCC 9637) and KO11 genomes using progressive Mauve.
E. coli W is at the top. Blocks of the same color are conserved in the two strains, and connected by lines. KO11 blocks are in the same orientation as in W if they appear above the black line, and opposite if below. White areas indicate regions that are present in only one strain. The tandem repeat region in KO11 is shown by the black bar below the white area, and IS10s in the repeat are shown by a single red triangle. IS10 insertions not associated with chromosomal rearrangements are shown by black triangles. The IS10 insertions at block junctions associated with inversions and deletions are numbered and shown by blue arrows. One inversion was produced by recombination between Mu prophage sequences (indicated by "Mu") but all other rearrangements involved IS10 elements.

The *E. coli W* genes absent from KO11 were placed in functional groups, and displayed as a pie chart (Fig. 6). Prophage genes were preferentially deleted in KO11.

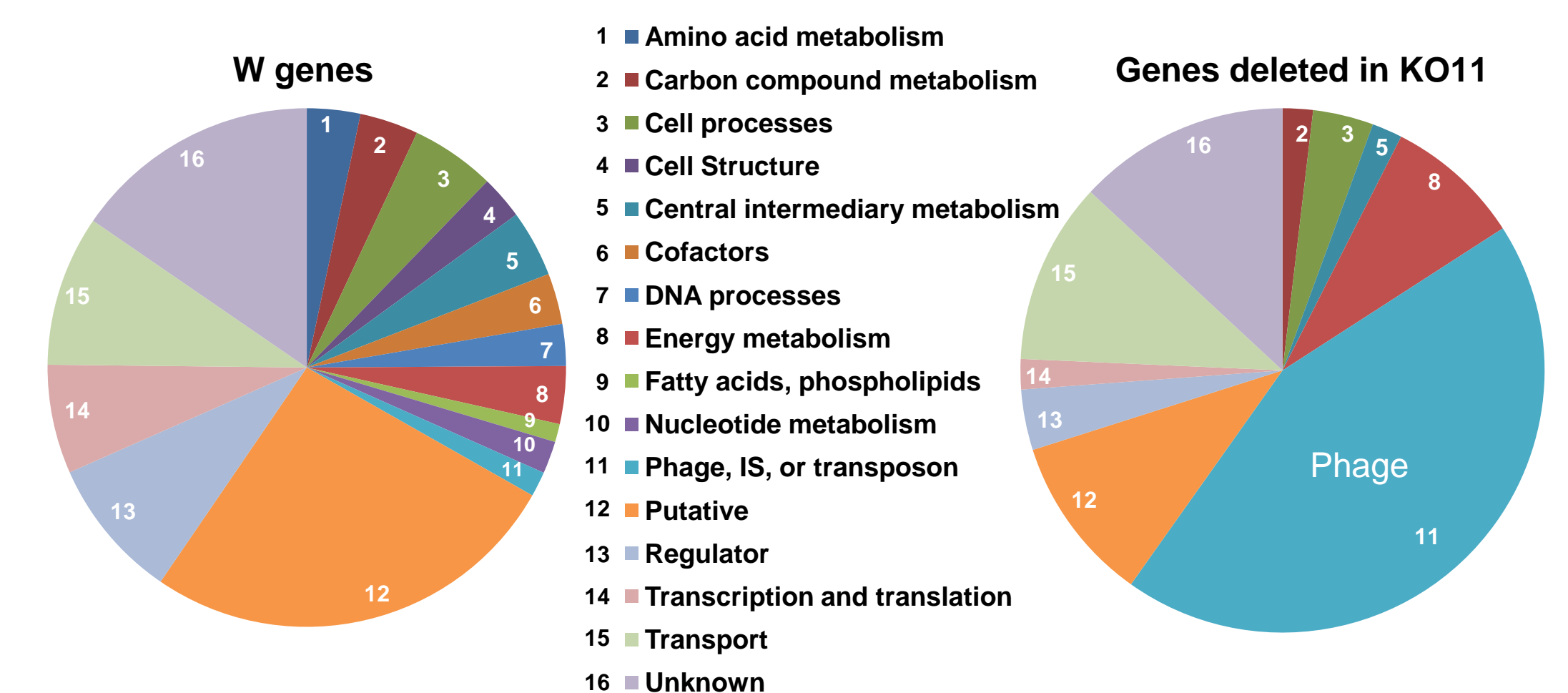


Figure 6. Functional groups for all W genes and for genes deleted in KO11.

A similar analysis was applied to genes that were altered in KO11 either by containing Single Nucleotide Polymorphisms (SNPs) or insertion of IS10 (Fig. 7). Genes involved in transport appear to contain more mutations than would be expected if the mutations were distributed over the functional groups at random.

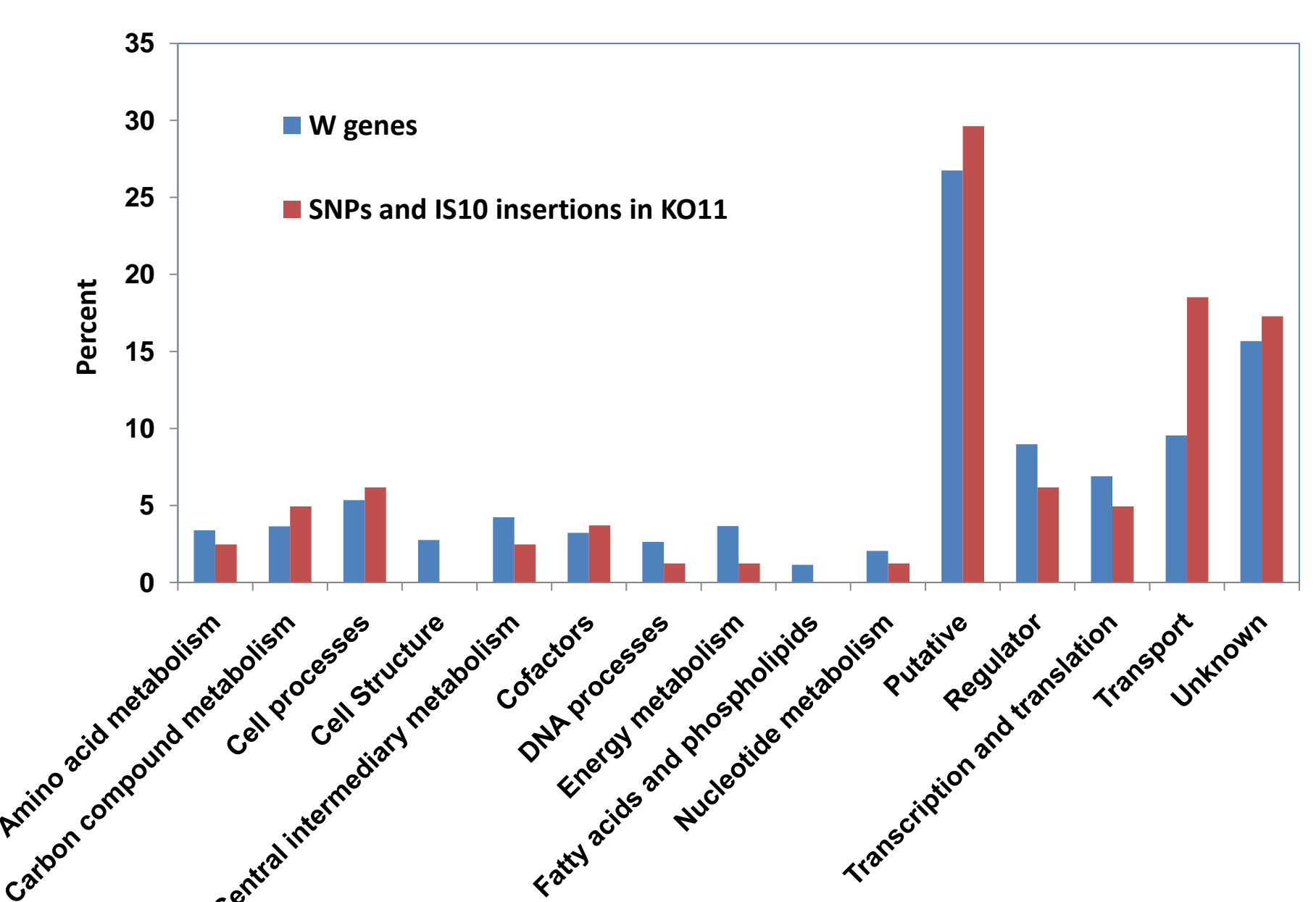


Figure 7. Functional groups for W genes and for genes altered in KO11.

CONCLUSIONS

- Completely assembled and annotated genomic sequences were obtained for *E. coli W* and KO11
- *E. coli W* is closely related to group B1 commensal strains
- KO11 contains a tandem duplication of ~20 copies of the *Z. mobilis pdc* and *adhB* genes
- KO11 contains IS10 at 30 different sites, and has chromosomal rearrangements relative to the parent W

RESULTS

The sequence of *E. coli W* was assembled from 454 sequence data using an optical restriction map to guide contig assembly. The W genome spans 4,897,452 bases and encodes 4,680 proteins, of which 801 are not present in *E. coli K-12*. In addition *E. coli W* contains plasmids pRK1 (102.5 kb) and pRK2 (5.4 kb).

REFERENCES

1. Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO. (2010) Metabolic engineering for production of bio-renewable fuels and chemicals: contributions of synthetic biology. *J Biomed Biotechnol.* 2010:761042.
2. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK. (2011) The genome sequence of *E. coli W* (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics* 12: 9.
3. Lucas S, Copeland A, Lapidus A, Cheng J-F, Bruce D, Goodwin L, Pittluck S, Land ML, Hauser L, Chang Y-J, Jeffries C, Tremaine M, Landick R, Keating D and Woyke TJ. (2010) *Escherichia coli W*, whole genome shotgun sequence. Genbank accession NZ_AEDF00000000.
4. Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO. (1991) Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl Environ Microbiol.* 57: 893-900.
5. Lucas S, Copeland A, Lapidus A, Cheng J-F, Goodwin L, Pittluck S, Munk AC, Dettler JC, Han C, Tapia R, Land M, Hauser L, Kyrpides N, Ivanova N, Ovchinnikova G, Pagani I, Keating D, Landick R and Woyke, T. (2011) *Escherichia coli* KO11, complete genome. Genbank accession CP002516.1.
6. Darling AE, Mau R, and Perna NT. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. *PLoS One.* 5(6):e11147.