

# Characterization of *Salmonella enterica* Subtypes by Optical Mapping

Colin W Dykes

OpGen Inc. 708 Quince Orchard Road, Gaithersburg, MD 20878



## Background

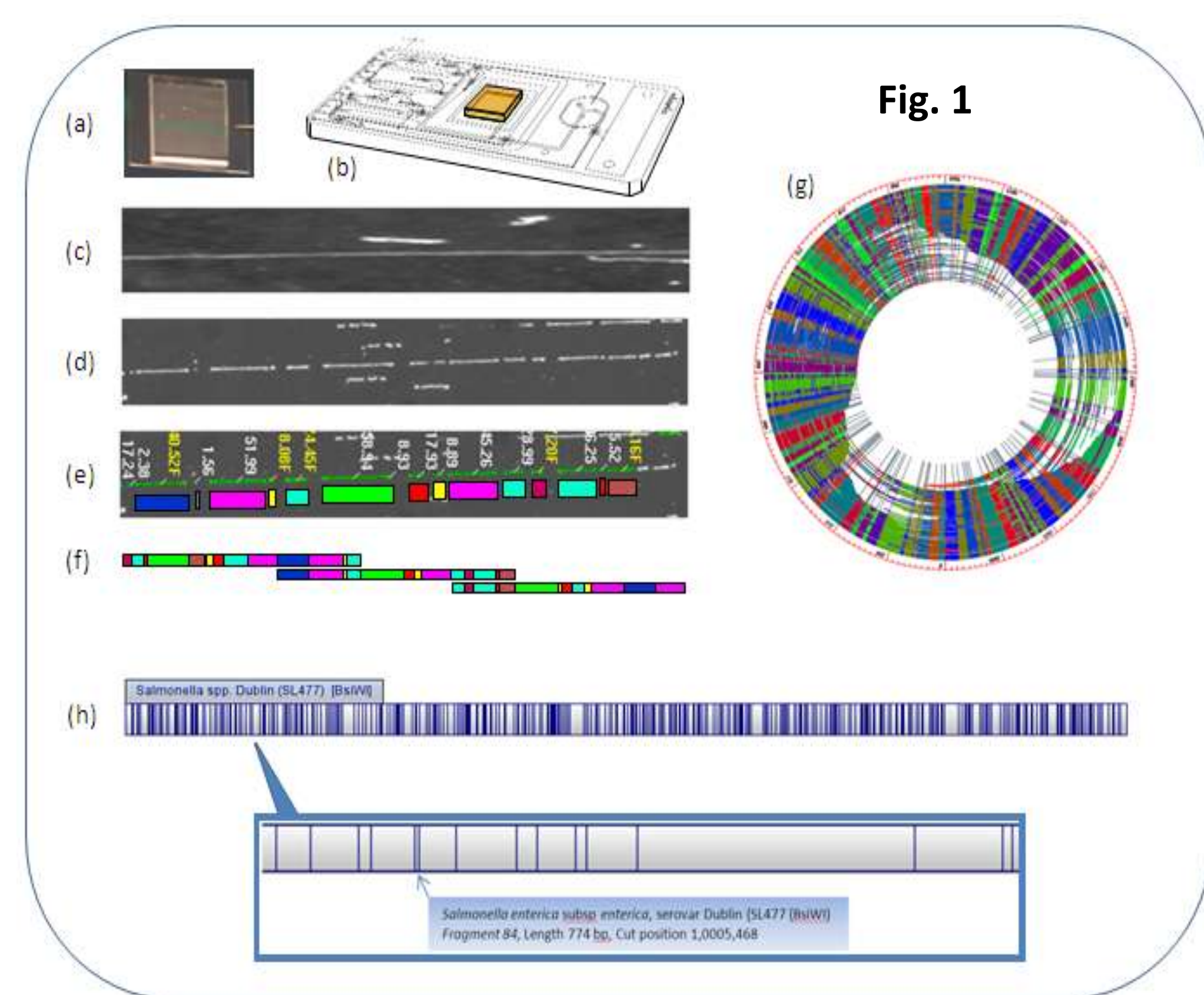
There is a pressing need for new rapid molecular methods for identification and sub-typing of *Salmonella* strains for surveillance, and for analysis of outbreaks. Optical Mapping is a method for producing ordered restriction maps across entire microbial genomes. For a typical 4 million base-pair bacterial genome, such maps normally contain around 400-600 contiguous fragments in the same relative order as found in the native chromosome. Comparison of Optical Maps from different isolates allows the detection of chromosomal rearrangements including insertions, deletions, inversions, and translocations (1). The latest automated systems have produced high-coverage Optical Maps from enriched bacterial cultures within 8 hours and require neither isolation of single colonies, nor pure cultures, as starting material.

The resolution afforded by Optical Mapping has been used to identify chromosomal markers in strains of *Escherichia coli* O157:H7 implicated in the 2006 “spinach” food-poisoning outbreak in the USA (2). The technology identified over a dozen chromosomal markers that distinguished the outbreak strain from other strains of O157:H7, mainly in prophages (1,2).

In this study, Optical Mapping was used to compare genetic diversity within, and between, reference *Salmonella* serovars.

## Methods

**Preparation of Optical Maps:** Chromosomal DNA from the *Salmonella* strains was subjected to analysis by Optical Mapping, as described previously (3). Briefly, genomic DNA from each isolate was captured in parallel arrays of long (>250 kb) chromosomal fragments on positively charged glass using a PDMS microfluidic device [Fig. 1(a)]. A microfluidic pumped cartridge device (b) is used to control digestion of the immobilized chromosomal fragments (c) with a restriction endonuclease (e.g. NcoI), to reveal cut restriction sites as “gaps” in the DNA by fluorescence microscopy after staining with the intercalating Dye JoJo-1 (d). The contiguous immobilized restriction fragments were sized by image analysis software that measures the fluorescence associated with each fragment and converts the optical images to ideograms where restriction fragments are represented by colored rectangles (e). Overlapping restriction fragment patterns in different molecules



were used to produce assemblies of molecules giving a minimum of 30X coverage over any part of the genome (f, g). The average size of each restriction fragment (measured in 30-100 different molecules in the assembly) was determined and used to create a linear “consensus map” (h) where each restriction site is represented by a vertical line in the horizontal rectangle. Sequence-based “*in silico*” whole genome restriction maps were also generated from sequenced reference genomes for alignment comparisons.

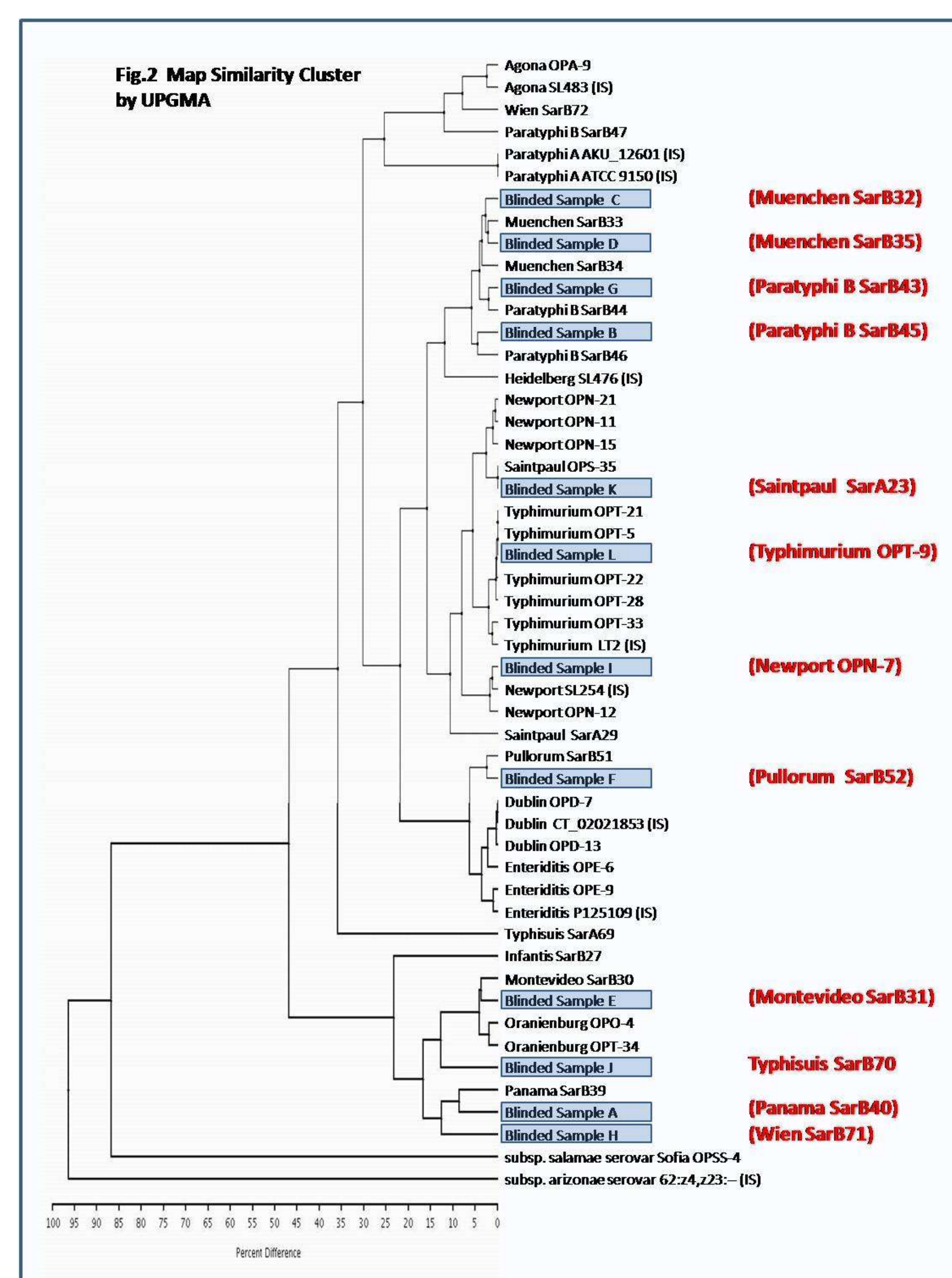
**Strain Clustering:** To construct a similarity cluster, maps were aligned using a dynamic programming algorithm based upon previous methods (4). This method finds the optimal alignment of two restriction maps according to a scoring model that incorporates fragment sizing errors, false and missing cuts, and missing small fragments. For a given alignment, the score is proportional to the log of the length of the alignment, penalized by the differences between the two maps, such that longer, better-matching alignments will have higher scores. From these alignments, a dissimilarity score for a pair of maps is calculated by adding up the lengths of the unmatched regions from both maps, and dividing this by the sum of the lengths of both maps in the pair. A matrix of these pair-wise scores is used as input to “agnes”, an agglomerative

clustering method implemented in the R statistical package, which creates dendrograms using UPGMA (unweighted pair group method with arithmetic mean, ref 5.)

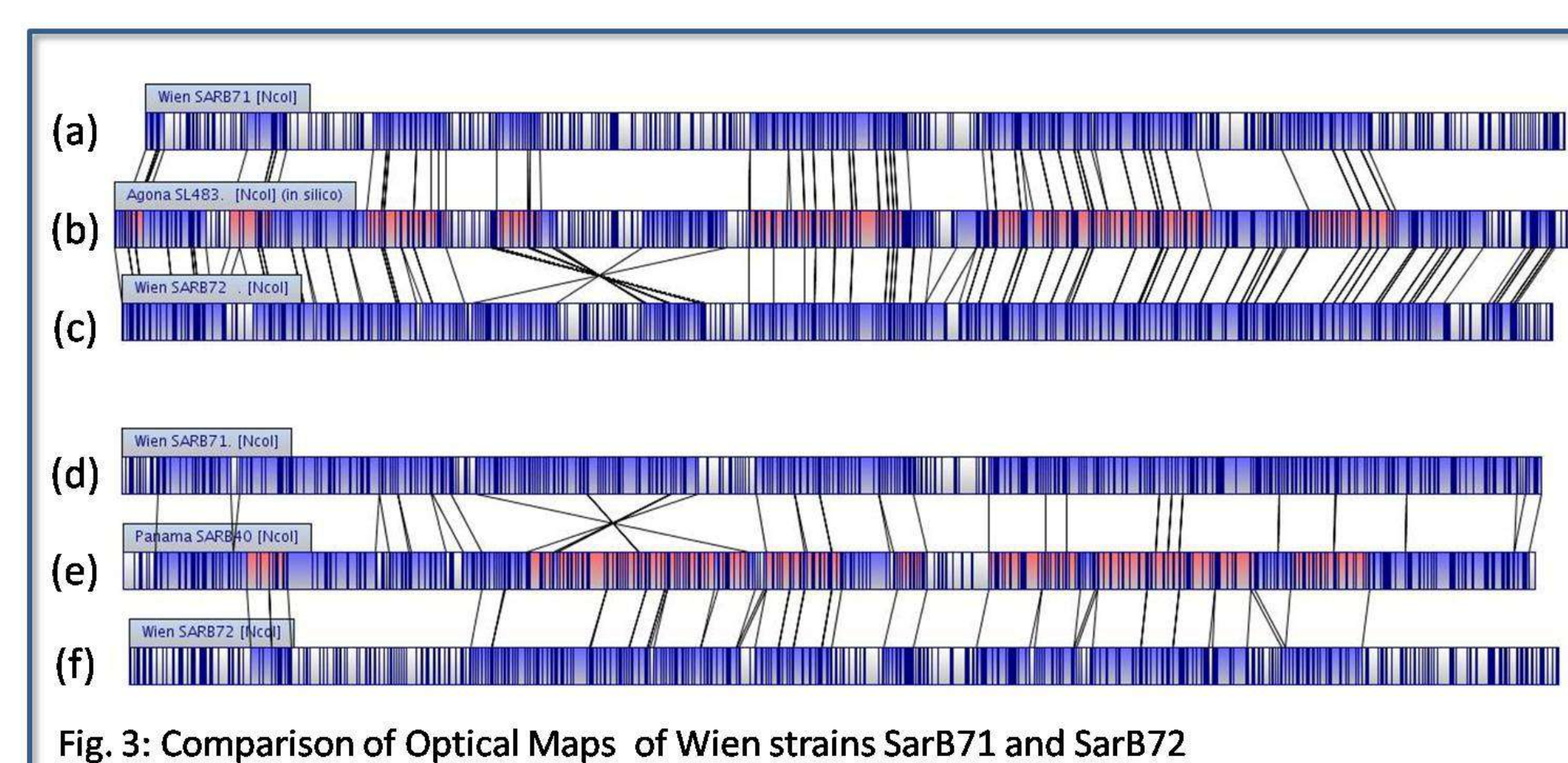
**Bacterial Strains:** Unless otherwise stated, all strains were *Salmonella enterica* subsp. *enterica*. For convenience, only the serovar names and isolate names are used here. Strains beginning with “SarA” or “SarB” were obtained from the University of Calgary *Salmonella* Genetic Stock Center. *In silico* Maps, i.e. Maps from sequenced strains, are designated (IS). “OP” strains are from the OpGen database. “Blinded samples” were selected primarily from the SarB reference collection and were identified by similarity to strains in the OpGen database using the UPGMA algorithm.

## Results

**Strain Clustering:** Fig.2 shows the relationship between *Salmonella* strains based on Optical Map comparison. In general, members of particular serovars showed a strong tendency to appear in the same group as other members of the same serovar, demonstrating a surprisingly strong correlation between chromosomal similarity and expressed antigens. Members of different subspecies (salamae, and arizonae) were strikingly different from each other, and from the subsp. *enterica* strains, at the Optical Map level.

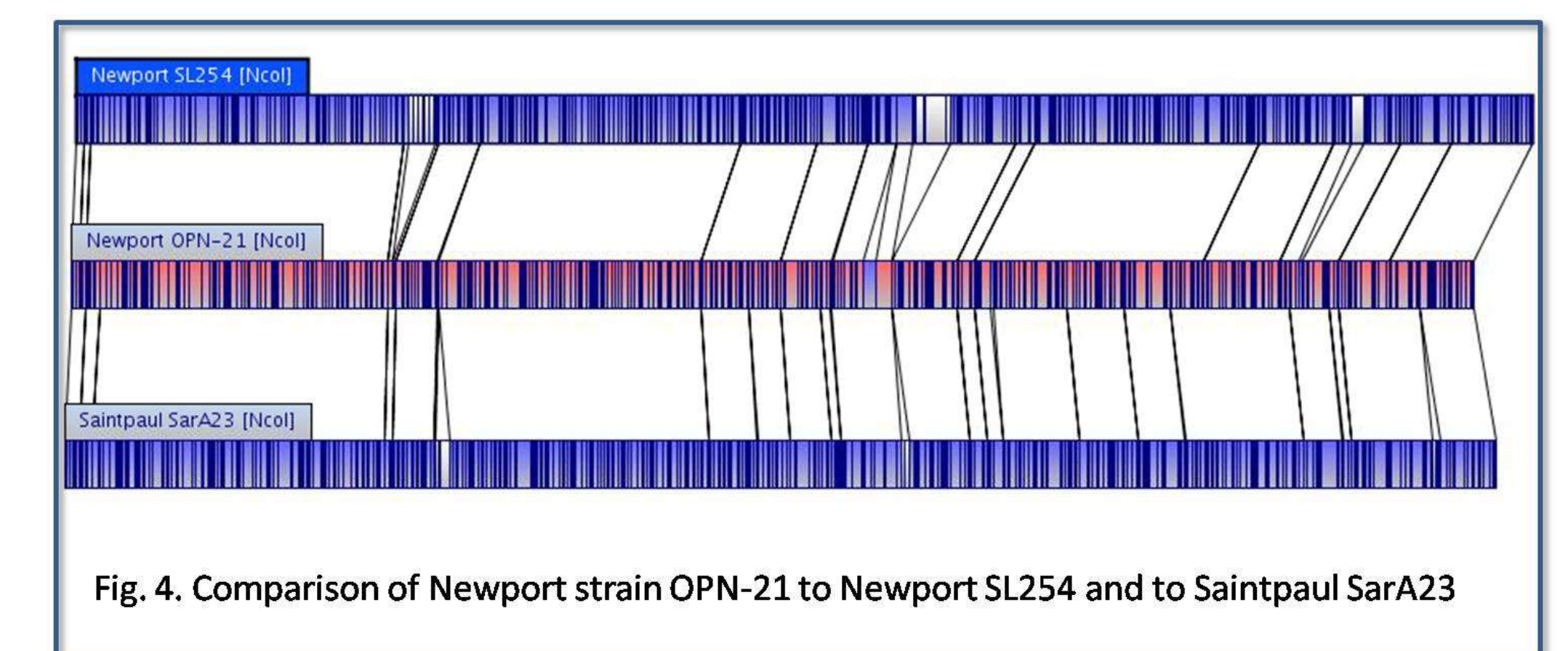


Strain clustering also demonstrated clear genomic dissimilarities between members of the same serovar and strong genomic similarities between members of different serovars. Optical Mapping not only detects these differences, but also allows them to be displayed in an intuitive manner using Mapsolver™ software, which highlights similar genomic regions in blue and dissimilar regions in white. Fig 3 shows that, as predicted from the clustering analysis,

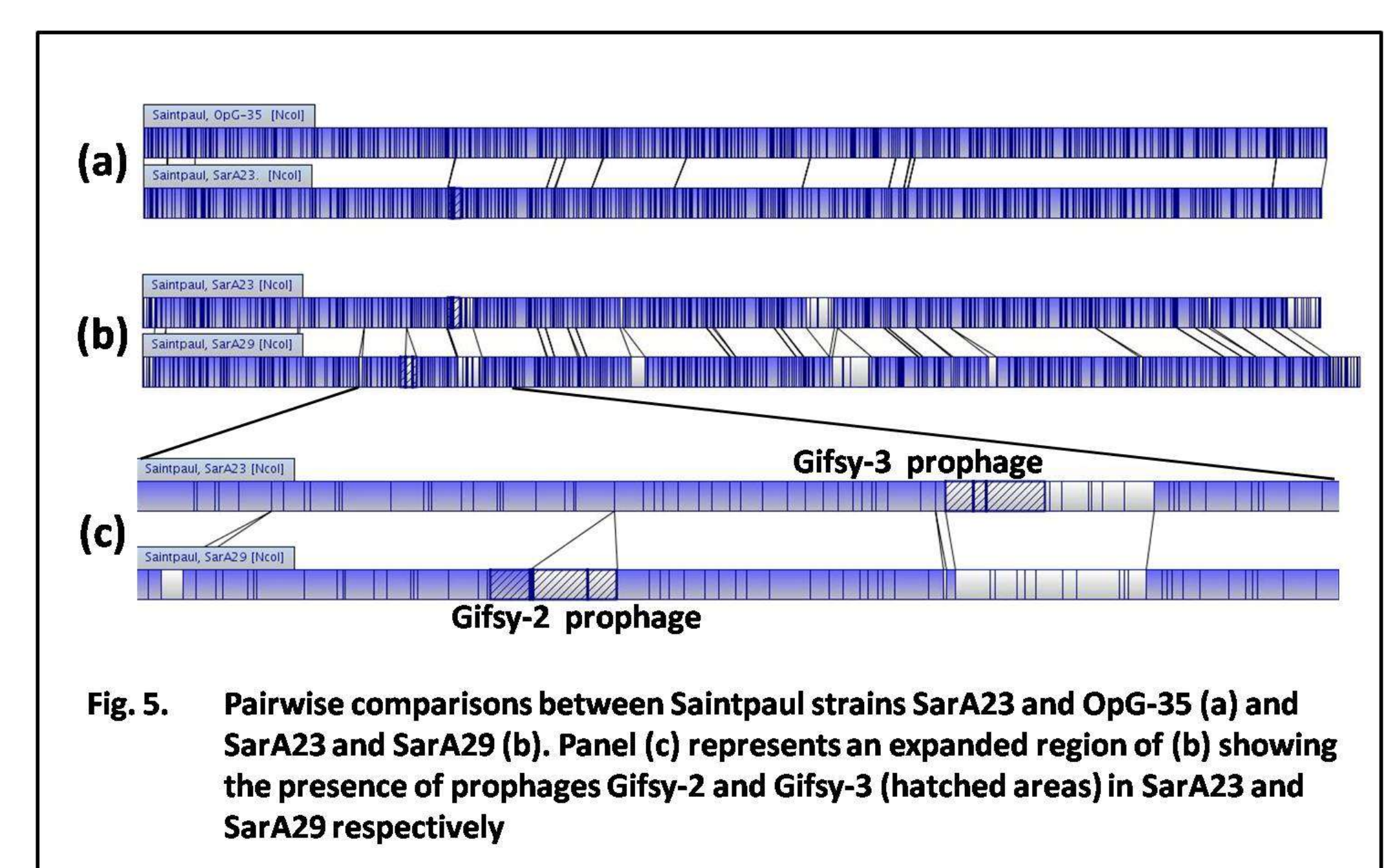


Wien strains SarB71 and SarB72 are quite dissimilar; SarB72 being much more similar to Agona strain SL483 than is Wien strain SarB71, as evidenced by the extent of genomic similarity (blue vs white) shown in Fig 3 (a) and (c). Regions shown in red are similar between all three strains. Conversely, Wien strain SarB71 is more closely related to the Panama strain SarB40 than is Wien strain SarB72. Mapsolver™ also indicates the nature and location of the genomic differences between these strains and shows the presence

of a large genomic inversion (indicated by “X” pattern in Fig. 3). Similarly, Newport strain OPN-21 is more closely related to Saintpaul strain SarA23 than it is to Newport SL254 (Fig. 4). Since Newport SL254 is a sequenced strain, the associated annotation can



be used to identify the genes and functions that are present in SL254 but not in OPN-21, adding more information to the analysis than is provided by other typing methods.



For example, comparison of Optical Maps of the closely related Saintpaul strains SarA23 and OpN-35, and the more distant SarA29 strain, with other sequenced strains, allows identification of mobile genetic elements such as Gifsy-2 and Gifsy-3 as contributing to the differences between SarA23 and SarA29 (Fig. 5).

## Conclusion

Optical Mapping can distinguish between closely-related *Salmonella* strains based on the ability to detect subtle genomic differences, including “indels” that may be due to the presence or absence of specific prophages, as well as other rearrangements such as inversions and translocations that are difficult to detect by other methods. This allows discrimination between different isolates of the same serovar and identification of the relatedness between genetically-similar strains belonging to different serovars. Comparison of Optical Maps of new isolates with *in silico* Optical maps of characterized strains can identify potential functional, or phenotypic differences between strains. The ability to detect even very small indels, with no apparent function, generates empirical markers to distinguish between otherwise identical isolates, providing a powerful tool for epidemiological studies.

## References

- Kotewicz ML, Mammel MK, LeClerc JE, Cebula TA (2008): Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* 154:3518-28.
- Kotewicz ML, Jackson SA, LeClerc JE, Cebula TA (2007): Optical maps distinguish individual strains of *Escherichia coli* O157:H7. *Microbiology* 153:1720-33.
- Cai W, Jing J, Irvin B, Ohler L, Rose E, Shizuya H, Kim UJ, Simon M, Anantharaman T, Mishra B, Schwartz DC (1998): High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. *PNAS* 95:3390-3395.
- Waterman, M. S., Smith, T. F. & Katcher, H. L. (1984). Algorithms for restriction map comparisons. *Nucleic Acids Res* 12, 237-242.
- Michener, C.D., Sokal, R.R. (1957): A quantitative approach to a problem of classification. *Evolution*, 11:490-499

**Footnote:** The data reported here were generated using a “manual” Optical Mapping system. OpGen now operates an automated “production” system that can generate fully-assembled Optical Maps within 8 hours of receipt of *Salmonella* cultures, and plans to reduce the time to Map to 3 hours from receipt of a culture.