

Problem

To evaluate whether *Escherichia coli* O26 isolates originating from humans and cattle can be assigned to distinct groups.

Background and Methods

DNA from 187 *E. coli* O26 isolates originating from humans and cattle were digested with the restriction enzyme *Xba* I and fragments separated using pulse-field gel electrophoresis (PFGE). Differences in DNA fragment length profile between isolates were used to construct a Jaccard similarity coefficient matrix using BioNumerics v3.0 assuming a tolerance of 1.3% and optimisation of 1.0%. The Jaccard similarity coefficient is the ratio of the number of DNA fragments of the same length to the number of DNA fragment-length positions between each pair of isolates and was used because it is simpler to interpret.

Data were further analysed using SAS v8.2. Similarities were converted into a data-based distance matrix and used to construct a dendrogram by the unweighted pair group mean average (UPGMA) method (*University of Kansas Science Bulletin*, **38**:1409-1438). UPGMA was used because:

- PFGE bands are neither homologous or independent. The use of character-based methods to build dendrograms and the evaluation of support for clusters using the bootstrap are not appropriate.
- The lack of information on a suitable outgroup means that there is no advantage to building a dendrogram using neighbour-joining.

Differences between human and cattle isolates were therefore evaluated by assigning isolates to statistically significant groups. Group membership was initially determined by estimating an optimum cut-off value using dendrogram-based distances, and isolates less separated by less than this value (0.393) assigned to the same group. The optimum cut-off value was calculated by:

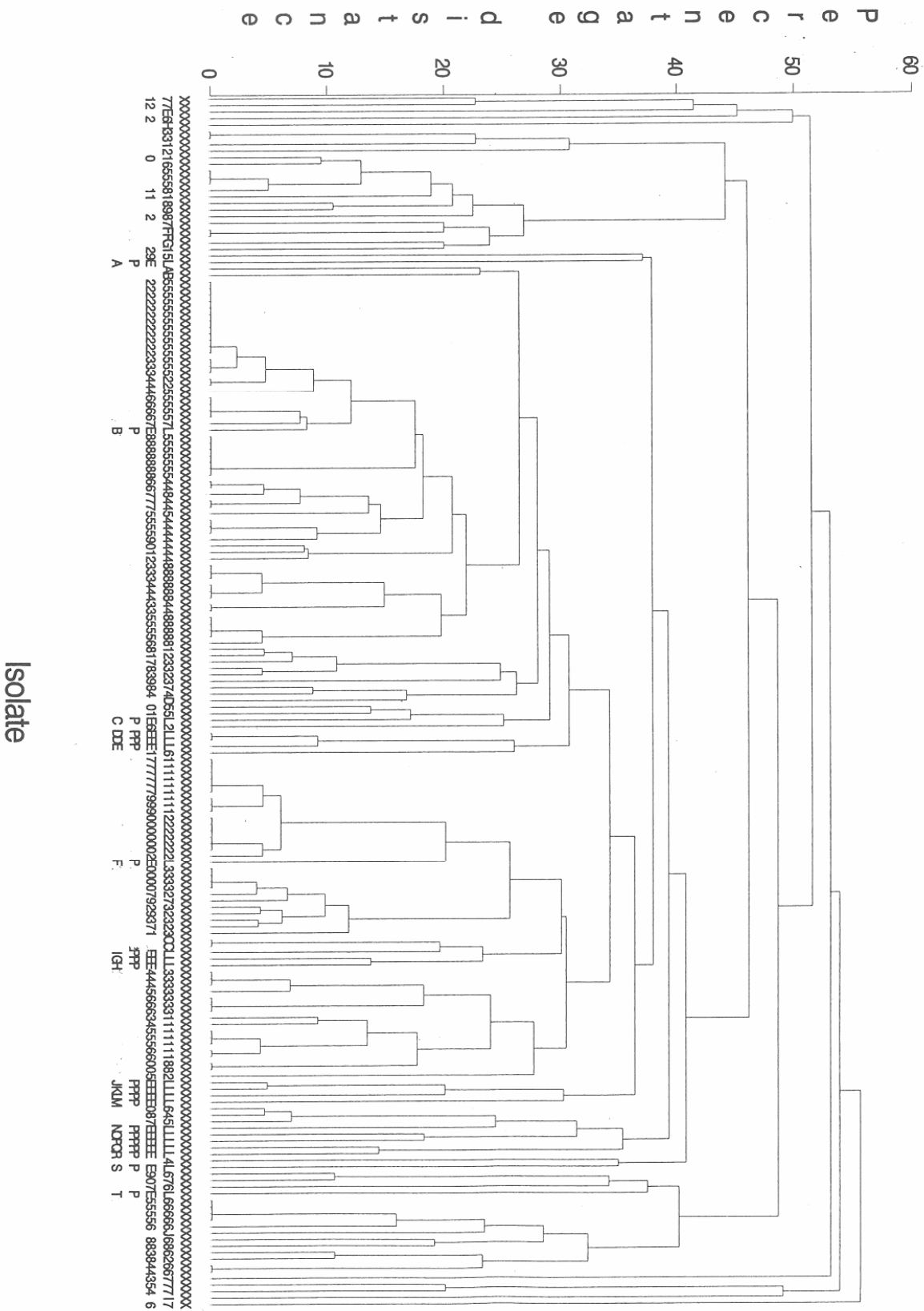
1. randomly choosing a value of between 0.00 and 1.00;
2. treating isolates within the groups identified by the cut-off value as identical (with a distance of 0) and isolates between groups as being entirely different (with a distance of 1). A group-based distance matrix was consisting of zeros and ones was constructed;
3. calculating the correlation coefficient between the comparable elements of the group and data-based distance matrices;
4. repeating 1000 times and selecting the cut-off value associated with the highest correlation coefficient.

Statistical support for each group was evaluated using the co-phenetic correlation. This is the correlation coefficient between the comparable elements of the data and dendrogram-based distance matrices. The statistical significance of co-phenetic correlations was evaluated using a Mantel test (*Cancer Research*, **27**:209-220) in which the probability of the estimate occurring by chance is estimated by repeatedly randomising the elements in one of the matrices and enumerating how often it occurs by chance. Programme *zt* (*Journal of Statistical Software*, **7**(10):1-12) was used for this purpose using 1000 simulations, and groups regarded as statistically significant when the probability of the co-phenetic correlation occurring by chance was 5% or less. Where a putative group was not statistically significant, a node forming a group with a higher distance was used, providing this did not subsume pre-existing statistically significant groups.

Results

- Data-based distances between isolates vary between 0 (i.e. isolates are identical) and 0.714.
- Dendrogram-based distances between isolates vary between 0.00 and 0.555
- The dendrogram for isolates is shown in Appendix 2a.
- Although there is some clustering of human isolates in the dendrogram (e.g. isolates XC, XLEPG, XLEPH & XLEPI), the closest neighbour of a human isolate is almost as often a cattle isolate (13 times) as a human isolate (16 times). Human isolates are interspersed amongst cattle isolates.
- Four groups were identified comprising 5, 19, 142 and 16 isolates. Average distances between these groups are 0.499, 0.442, 0.406 and 0.400 respectively. All four groups are supported by statistically significant co-phenetic correlations ranging from 0.807 to 0.945. Membership of the groups is given in Appendix 2b. Five isolates were not assigned to a group.
- Analysis using the Dice similarity coefficient result in the same group membership.
- All four groups include *E. coli* O26 isolates of human and cattle origin. This pattern is consistent with human isolates being interspersed with cattle isolates.

UPGMA dendrogram of isolates based on Jaccards distance



Appendix 3 – STATISTICAL REPORT ON PFGE ANALYSIS

Isolate membership for each group

Group5	Group10	Group12	Group13
X62	X1	X13	X28
X71	X10	X14	X63
X72	X11	X15.1	X64.1
XE	X12	X15.2	X64.2
XH	X2	X15.3	X65.1
	X3.1	X16.1	X65.2
	X3.2	X16.2	X65.3
	X5.1	X17.1	X65.4
	X5.2	X17.2	X66
	X5.3	X17.3	X67
	X6	X17.4	X68
	X7	X17.5	X69
	X8	X17.6	X70
	X81	X18	X88
	X82	X19.1	XJ
	X9	X19.2	XLEPT
	XF.1	X19.3	
	XF.2	X20.1	
	XG	X20.2	
		X20.3	
		X20.4	
		X20.5	
		X20.6	
		X21	
		X22	
		X23	
		X24.1	
		X24.2	
		X25	
		X26	
		X27.1	
		X27.2	
		X29	
		X30.1	
		X30.2	
		X30.3	
		X30.4	
		X31	
		X32	

Appendix 3 – STATISTICAL REPORT ON PFGE ANALYSIS

Group5	Group10	Group12	Group13
		X33	
		X34.1	
		X34.2	
		X34.3	
		X35	
		X36.1	
		X36.2	
		X36.3	
		X37	
		X38	
		X39	
		X4	
		X40	
		X41	
		X42	
		X43.1	
		X43.2	
		X44	
		X45.1	
		X45.2	
		X45.3	
		X46.1	
		X46.2	
		X47.1	
		X47.2	
		X48	
		X49	
		X50	
		X51	
		X52.1	
		X52.10	
		X52.11	
		X52.12	
		X52.2	
		X52.3	
		X52.4	
		X52.5	
		X52.6	
		X52.7	
		X52.8	
		X52.9	

Appendix 3 – STATISTICAL REPORT ON PFGE ANALYSIS

Group5	Group10	Group12	Group13
		X53.1	
		X53.2	
		X53.3	
		X54	
		X55	
		X56.1	
		X56.2	
		X56.3	
		X56.4	
		X57	
		X58.1	
		X58.2	
		X58.3	
		X58.4	
		X58.5	
		X58.6	
		X58.7	
		X59	
		X60	
		X61	
		X77	
		X78	
		X79	
		X80.1	
		X80.2	
		X83.1	
		X83.2	
		X83.3	
		X84.1	
		X84.2	
		X84.3	
		X85.1	
		X85.2	
		X85.3	
		X85.4	
		X86	
		X87	
		XA	
		XB	
		XC.1	
		XC.2	

Appendix 3 – STATISTICAL REPORT ON PFGE ANALYSIS

Group5	Group10	Group12	Group13
		XD	
		XLEPA	
		XLEPB	
		XLEPC	
		XLEPD.1	
		XLEPD.2	
		XLEPE	
		XLEPF	
		XLEPG	
		XLEPH	
		XLEPI	
		XLEPJ	
		XLEPK	
		XLEPL	
		XLEPM	
		XLEPN	
		XLEPO	
		XLEPP	
		XLEPQ	
		XLEPR	
		XLEPS	