Watts MJ. 2011. Using data clustering as a method of estimating the risk of establishment of bacterial crop diseases. Computational Ecology and Software, 1(1): 1-13

Supplementary Material

Table S.1 Results of null-model analysis using random knight's tour algorithm.

Measure	Observed	Simulated	p(o<= e)	p(o>= e)
C-score	817.329	760.7315/0.71086	1.0	0.0
Checkerboard	539.0	622.645/1947.883	0.024	0.976
Combinations	364.0	370.377/3.90277	0.001	1.0

Table S.2 SOM output map dimensions and training epochs.

X-dimension	Y-dimension	#neurons	epochs
2	4	8	4000
3	5	15	7500
4	6	24	12000
5	7	35	17500
6	9	54	27000
7	11	77	38500
8	12	96	48000
9	12	108	54000

Table S.3 Top eighty species by risk weighting, as determined by Kohonen SOM clusters. 'Species' is the scientific name of the species. 'Risk Weighting' presents the mean and standard deviation of the risk weighting for that species over all 100 trials. 'Presence' indicates the species' presence (p) or absence (a) in the target region (New Zealand).

Species	**Risk Weighting**	**Presence**
Rhizobium radiobacter	1/0	p
Clavibacter michiganensis subsp. michiganensis	1/0	p
Pseudomonas savastanoi pv. phaseolicola	1/0	p
Pseudomonas syringae pv. pisi	1/0	p
Xanthomonas axonopodis pv. phaseoli	1/0	p
Xanthomonas vesicatoria	1/0	p
Erwinia carotovora subsp. carotovora	0.89/0	p
Burkholderia andropogonis	0.89/0	p
Pseudomonas savastanoi pv. glycinea	0.89/0	p
Pseudomonas syringae pv. lachrymans	0.89/0	p
Pseudomonas syringae pv. tabaci	0.89/0	p
Xanthomonas arboricola pv. pruni	0.89/0	p
Clavibacter michiganensis subsp. insidiosus	0.78/0	p
Erwinia chrysanthemi	0.78/0	p
Pseudomonas syringae pv. maculicola	0.78/0	p
Pseudomonas syringae pv. syringae	0.78/0	p
Xanthomonas campestris pv. campestris	0.78/0	p
Xanthomonas arboricola pv. juglandis	0.78/0	p
aster yellows phytoplasma group	0.67/0	a
Erwinia carotovora subsp. atroseptica	0.67/0	p
Potato witches broom phytoplasma	0.67/0	a
Pseudomonas syringae pv. mori	0.67/0	p
Pseudomonas syringae pv. tomato	0.67/0	p
Xanthomonas axonopodis pv. glycines	0.67/0	a
Xanthomonas axonopodis pv. malvacearum	0.67/0	a
Xanthomonas hortorum pv. pelargonii	0.67/0	a
Rhizobium rhizogenes	0.56/0	a
Rhodococcus fascians	0.56/0	p
Erwinia amylovora	0.56/0	p
Pseudomonas syringae pv. coronafaciens	0.56/0	p
Ralstonia solanacearum race 1	0.56/0	p
Bacillus subtilis	0.44/0	a

clover phyllody phytoplasma	0.44/0	a
Curtobacterium flaccumfaciens pv. flaccumfaciens	0.44/0	a
Pseudomonas syringae pv. atrofaciens	0.44/0	p
Pseudomonas cichorii	0.44/0	p
Pseudomonas marginalis pv. marginalis	0.44/0	p
Pseudomonas syringae pv. striafaciens	0.44/0	a
Xanthomonas arboricola pv. corylina	0.44/0	a
Apple rubbery wood phytoplasma	0.33/0	p
Rathayibacter rathayi	0.33/0	p
Clavibacter michiganensis subsp. sepedonicus	0.33/0	a
Pectobacterium chrysanthemi	0.33/0	p
Erwinia tracheiphila	0.33/0	a
Erwinia chrysanthemi pv. zeae	0.33/0	p
pear decline phytoplasma	0.33/0	a
Pseudomonas corrugata	0.33/0	p
Burkholderia gladioli pv. gladioli	0.33/0	a
Pseudomonas syringae pv. morsprunorum	0.33/0	a
Pseudomonas savastanoi pv. savastanoi	0.33/0	p
Streptomyces scabiei	0.33/0	a
Xanthomonas axonopodis pv. begoniae	0.33/0	p
Xanthomonas hortorum pv. carotae	0.33/0	a
Xanthomonas fragariae	0.33/0	a
Xanthomonas vasicola pv. holcicola	0.33/0	p
Xanthomonas hyacinthi	0.33/0	a
Xanthomonas translucens pv. translucens	0.33/0	a
apple proliferation phytoplasma	0.22/0	a
Clavibacter michiganensis subsp. nebraskensis	0.22/0	a
Curtobacterium flaccumfaciens pv. poinsettiae	0.22/0	p
Erwinia chrysanthemi pv. chrysanthemi	0.22/0	a
Enterobacter dissolvens	0.22/0	a
Pantoea stewartii	0.22/0	a
peach X-disease phytoplasma	0.22/0	a
peach yellows phytoplasma	0.22/0	a
Acidovorax avenae subsp. avenae	0.22/0	a
Burkholderia cepacia	0.22/0	a
Pseudomonas syringae pv. delphinii	0.22/0	p
Pseudomonas fuscovaginae	0.22/0	a

Pseudomonas syringae pv. helianthi	0.22/0	p
Pseudomonas rubrisubalbicans	0.22/0	p
Ralstonia solanacearum race 3	0.22/0	a
Pseudomonas tolaasii	0.22/0	p
Rubus stunt phytoplasma	0.22/0	a
Xanthomonas axonopodis pv. dieffenbachiae	0.22/0	a
Xanthomonas hortorum pv. hederae	0.22/0	p
Xylella fastidiosa	0.22/0	a
Strawberry witches broom phytoplasma	0.11/0	a
Leifsonia xyli subsp. xyli	0.11/0	a
Rathayibacter tritici	0.11/0	a

Table S.4 Top eighty species by risk weighting, as determined by k-means. Column headings are as for Table S.3.

Species	Risk Weighting	Presence
Rhizobium radiobacter	0.99/0.05	p
Pseudomonas syringae pv. syringae	0.93/0.08	p
Pseudomonas savastanoi pv. phaseolicola	0.91/0.12	p
Xanthomonas arboricola pv. juglandis	0.88/0.11	p
Clavibacter michiganensis subsp. michiganensis	0.88/0.13	p
Xanthomonas campestris pv. campestris	0.86/0.13	p
Pseudomonas syringae pv. lachrymans	0.86/0.11	p
Erwinia carotovora subsp. atroseptica	0.86/0.14	p
Erwinia chrysanthemi	0.85/0.15	p
Erwinia carotovora subsp. carotovora	0.83/0.13	p
Pseudomonas syringae pv. pisi	0.79/0.16	p
Xanthomonas axonopodis pv. phaseoli	0.78/0.15	p
Pseudomonas syringae pv. tabaci	0.74/0.19	p
Xanthomonas vesicatoria	0.73/0.2	p
Erwinia amylovora	0.75/0.26	p
Pseudomonas syringae pv. tomato	0.69/0.14	p
Pseudomonas marginalis pv. marginalis	0.66/0.19	p
Pseudomonas savastanoi pv. glycinea	0.66/0.22	p
Xanthomonas hortorum pv. pelargonii	0.66/0.16	a
Pseudomonas syringae pv. maculicola	0.67/0.22	p
Apple rubbery wood phytoplasma	0.65/0.21	p
Pseudomonas syringae pv. mori	0.64/0.18	p
Pseudomonas cichorii	0.6/0.21	p
Pseudomonas savastanoi pv. savastanoi	0.61/0.15	p
Rhodococcus fascians	0.58/0.24	p
Pseudomonas viridiflava	0.57/0.21	p
Pseudomonas syringae pv. morsprunorum	0.58/0.21	a
Erwinia chrysanthemi pv. zeae	0.56/0.21	p
Xanthomonas arboricola pv. pruni	0.55/0.27	p
Pseudomonas corrugata	0.53/0.2	p
Clavibacter michiganensis subsp. insidiosus	0.53/0.23	p
Xanthomonas axonopodis pv. begoniae	0.52/0.19	p
Pseudomonas syringae pv. coronafaciens	0.48/0.15	p
Pectobacterium chrysanthemi	0.52/0.28	p

aster yellows phytoplasma group	0.47/0.18	a
Xanthomonas axonopodis pv. malvacearum	0.49/0.26	a
pear decline phytoplasma	0.5/0.27	a
Pseudomonas tolaasii	0.47/0.19	p
Burkholderia andropogonis	0.47/0.26	p
Ralstonia solanacearum race 3	0.45/0.17	a

Rathayibacter rathayi	0.44/0.21	p
Streptomyces scabiei	0.43/0.18	a
Ralstonia solanacearum race 1	0.44/0.28	p
Xanthomonas arboricola pv. corylina	0.4/0.16	a
Burkholderia gladioli pv. gladioli	0.41/0.2	a
Xanthomonas vasicola pv. holcicola	0.38/0.17	p
Xanthomonas axonopodis pv. glycines	0.38/0.22	a
Pseudomonas syringae pv. delphinii	0.39/0.2	p
clover phyllody phytoplasma	0.4/0.21	a
Bacillus subtilis	0.37/0.18	a
Xanthomonas hyacinthi	0.35/0.15	a
Pseudomonas syringae pv. atrofaciens	0.34/0.16	p
Potato witches broom phytoplasma	0.35/0.24	a
Rhizobium rhizogenes	0.35/0.23	a
Xanthomonas translucens pv. translucens	0.32/0.19	a
Spiroplasma citri	0.34/0.16	p
Pectobacterium rhapontici	0.34/0.16	a
apple proliferation phytoplasma	0.38/0.26	a
Xanthomonas fragariae	0.35/0.22	a
Pantoea agglomerans	0.34/0.17	p
Xanthomonas hortorum pv. carotae	0.3/0.21	a
Pseudomonas syringae pv. aptata	0.29/0.13	p
Xanthomonas hortorum pv. hederae	0.32/0.21	p
Acidovorax avenae subsp. avenae	0.29/0.26	a
Clavibacter michiganensis subsp. sepedonicus	0.29/0.16	a
Pseudomonas syringae pv. helianthi	0.26/0.15	p
Rubus stunt phytoplasma	0.29/0.21	a
Erwinia chrysanthemi pv. chrysanthemi	0.26/0.17	a
Pseudomonas syringae	0.23/0.14	a

European stone fruit yellows phytoplasma	0.28/0.24	a
Pseudomonas syringae pv. striafaciens	0.22/0.18	a
Xanthomonas axonopodis pv. vitians	0.21/0.16	a
Grapevine yellows phytoplasmas	0.25/0.19	a
Xanthomonas axonopodis pv. dieffenbachiae	0.22/0.17	a
Xanthomonas populi	0.26/0.23	a
Xanthomonas campestris	0.19/0.14	a
Curtobacterium flaccumfaciens pv. flaccumfaciens	0.19/0.17	a
Pseudomonas fluorescens	0.18/0.12	a
Curtobacterium flaccumfaciens pv. poinsettiae	0.18/0.1	p
Leifsonia xyli subsp. xyli	0.2/0.22	a

Table S.5 Results of comparisons between risk weightings from SOM and k- means. 'Clusters' is the number of clusters or output map neurons used. 'Weight Diff.' is the Euclidean distance between the species risk weightings across all species. 'Neighbour Freqs. Diff.' is the Euclidean distance between the regional neighbour frequencies across all regions.

Clusters	**Weight Diff.**	**Neighbour Freqs. Diff.**
8	4.21	3.30
15	4.36	2.31
24	4.14	2.57
35	4.43	1.72
54	4.84	1.66
77	4.47	1.92
96	4.98	0.88
108	5.10	0.78

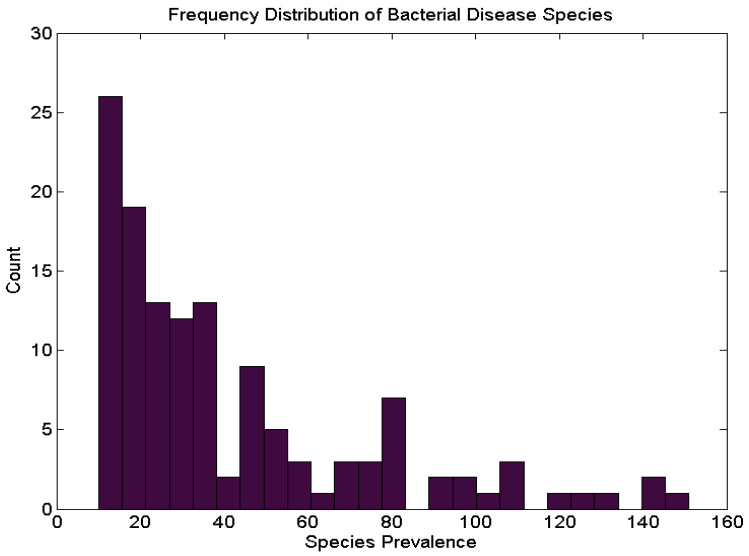

Fig. S.1 Frequency distribution of species prevalence.

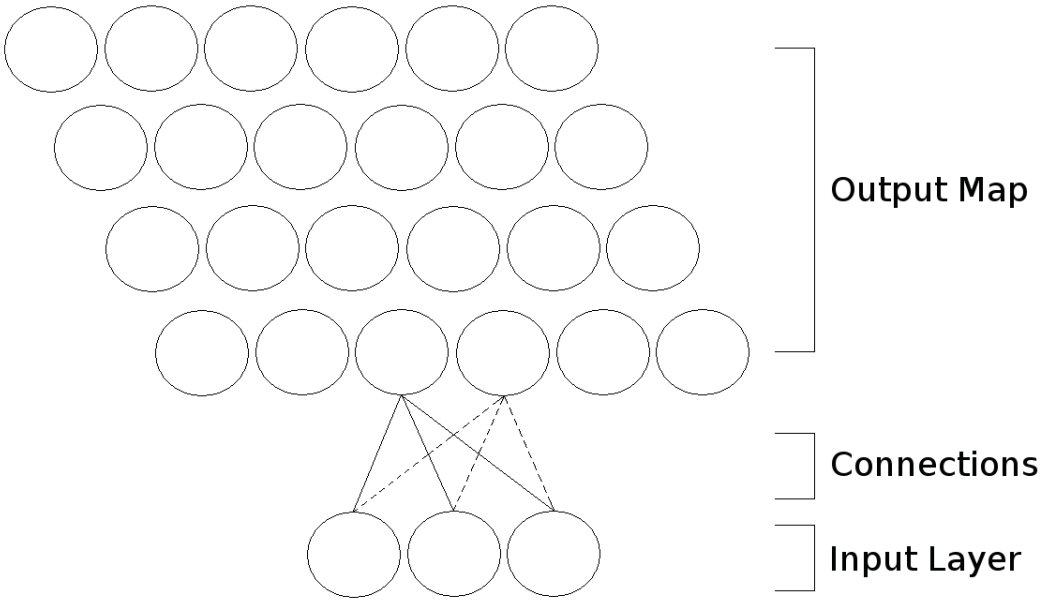

Fig. S.2 Structure of Kohonen Self-Organising Map. For clarity, only the connections to two output map neurons are shown.