A typing scheme for the honey bee pathogen *Melissococcus plutonius* allows detection of disease transmission events and a study of the distribution of variants

Edward Haynes[1,2], Thorunn Helgason[1], J. Peter W. Young[1], Richard Thwaites[2], Giles E. Budge[2]

Corresponding author;

Dr Thorunn Helgason

Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

Tel; +44 1904 328614

Fax; +44 1904 328505

Email; thorunn.helgason@york.ac.uk

Running Title: A typing scheme for *Melissococcus plutonius*

Citation:

Haynes, E., Helgason, T., Young, J. P. W., Thwaites, R. and Budge, G. E. (2013), A typing scheme for the honeybee pathogen *Melissococcus plutonius* allows detection of disease transmission events and a study of the distribution of variants. Environmental Microbiology Reports, 5: 525–529. doi:10.1111/1758-2229.12057

1) University of York, UK 2) The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK

20 ## Summary

21 *Melissococcus plutonius* is the bacterial pathogen that causes European Foulbrood of honey bees, a
22 globally important honey bee brood disease. We have used next generation sequencing to identify
23 highly polymorphic regions in an otherwise genetically homogenous organism, and used these loci to
24 create a modified MLST scheme. This synthesis of a proven typing scheme format with next
25 generation sequencing combines reliability and low costs with insights only available from high
26 throughput sequencing technologies. Using this scheme we show that the global distribution of *M.*
27 *plutonius* variants is not uniform. We use the scheme in epidemiological studies to trace movements
28 of infective material around England, insights that would have been impossible to confirm without
29 the typing scheme. We also demonstrate the persistence of local variants over time.

30 ## Introduction

31 European Foulbrood of honey bees (EFB) is an important disease of honey bee larvae, which
32 is found on every continent where bees are managed. Infected larvae become displaced in their
33 cells, covered in a waxy sheen before decomposing and discolouring (Bailey, 1981). Whilst not
34 always the case, the disease can prove fatal to a colony (Bailey, 1981). EFB is caused by the Gram-
35 positive bacterium *Melissococcus plutonius* (White, 1912; Bailey, 1957) which despite being globally
36 distributed has been found to exhibit extremely low levels of genetic diversity (Allen and Ball, 1993;
37 Djordjevic et al., 1999). Indeed, isolates from the UK and Australia have proven indistinguishable by
38 RFLP (Restriction Fragment Length Polymorphism). In spite of apparently low genetic diversity in the
39 pathogen population, some differences have been observed in biochemical and physical
40 characteristics (Allen and Ball, 1993). More recently a subtype of *M. plutonius*, showing genetic and
41 metabolic differences from previously reported *M. plutonius*, was reported in Japan (Arai et al.,
42 2012), suggesting *M. plutonius* may contain more genetic variation than previously thought.

43 Disease prevalence varies annually and regionally (Wilkins et al., 2007; Roetschi et al., 2008;
44 Budge et al., 2010). *M. plutonius* is able to persist in a honey bee without causing symptoms
45 (Forsgren et al., 2005; Roetschi et al., 2008; Budge et al., 2010), which may be exacerbated by the
46 presence of secondary bacteria (Bailey, 1981). Regional differences in prevalence are often found:
47 for example in the UK, in 2012, 1.9% of inspected apiaries in Wales were EFB-positive, compared
48 with 6.6% in England (NBU, 2012a). Recent evidence suggests that levels of EFB infection are
49 increasing in Switzerland and the UK (Wilkins et al., 2007; Roetschi et al., 2008; NBU, 2012b) and
50 countries thought to be disease-free have suffered recent outbreaks (e.g. Norway and Scotland).
51 Despite this growing risk, the epidemiology of EFB is not fully understood.

52 Multi locus sequence typing (MLST) has proved a successful and powerful way of
53 distinguishing pathogen variants and inferring an understanding of disease aetiology (Killgore et al.,
54 2008). A traditional MLST scheme uses sequence variation in six to ten housekeeping genes, with
55 each gene representing a different locus and any variation representing different alleles (Maiden et
56 al., 1998). Such a scheme uses housekeeping genes because they are conserved enough to be
57 present in all isolates, but should show enough variability to enable different variants to be
58 identified (Urwin and Maiden, 2003). Genetically distinct isolates can then be grouped into variants
59 or Sequence Types (ST) based on different allelic profiles across all loci. MLST schemes have been
60 developed which do not correspond exactly to the original concept of an MLST, as they contain non-
61 housekeeping gene loci (Ahmed et al., 2006), and the scheme developed here makes use of genomic
62 comparisons to identify appropriate loci when dealing with a genetically homogeneous bacterium.
63 The lack of a typing scheme for *M. plutonius* has severely limited epidemiological investigations into
64 this important disease.

65 We present the first usable typing scheme for *M. plutonius*, which is able to distinguish
66 multiple STs of the bacterium and to identify epidemiologically relevant transmission routes. We also
67 characterise the geographical distribution of STs.

68 ## Results and Discussion

69 ### Locus selection using traditional methods

70 Loci used in MLST schemes for the closely related *Enterococcus faecalis* (Ruiz-Garbajosa et
71 al., 2006) and *Enterococcus faecium* (Homan et al., 2002) were tested initially. In addition, some
72 genes encoding proteins with predicted transmembrane domains were tested, because proteins
73 with extracellular regions often show high diversity driven by interactions between pathogen and
74 host (Gupta and Maiden, 2001). We tested 15 such loci on a test set of 21 isolates. These isolates
75 included the Type Strain (LMG 20360) and 18 other epidemiologically unrelated isolates from a
76 broad geographic area across the UK, as well as single isolates from Thailand and Australia.

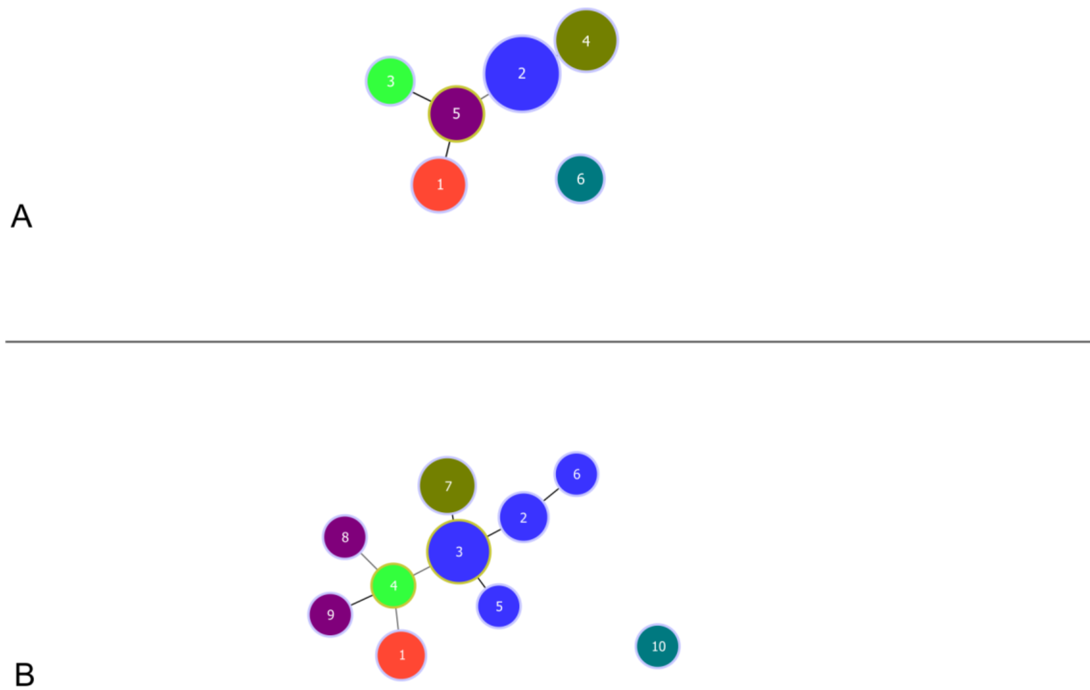
77 Two loci in the test set showed no variation on sequencing. Nine of the loci had two alleles
78 each, but the second allele was only present in isolate 7596, later shown to belong to the atypical
79 group. The remaining four loci each had three alleles. When all 15 loci were used, they split the test
80 set of isolates into six STs.

81 ### Locus selection using genomic comparison

82 Subsequently, four loci were identified by genomic comparisons among the Type Strain and
83 nine isolates sampled from the UK in 2010 (genomic data not shown, comparisons were performed
84 with the Mauve package (Darling et al., 2010)). These loci split the same set of 21 isolates into ten
85 STs. The 15 loci tested earlier provided no additional information, so the decision was made to
86 continue with the scheme using only the four loci identified using genomic comparison (Figure 1).

87

88



89

90 Figure 1. goeBURST (Francisco et al., 2009) trees of STs. Each circle represents a different ST, with
91 lines linking closest relatives. Black lines indicate a single allelic change between STs, light grey lines
92 indicate differences at two loci. Circles ringed with a yellow outline indicate putative founder
93 genotypes. 1A shows putative pattern of descent based upon the 15 traditional MLST loci. 1B shows
94 the putative pattern of descent based upon the four loci identified by genomic comparisons. In both
95 1A and 1B the numbers in the circles show the ST as defined by the four locus scheme, and the
96 colours show the STs as defined by the 15 traditional loci – that is if a circle has the same colours in
97 1A and 1B, they contain the same isolates. It is clear to see how groups identified by the 15
98 traditional loci have been subdivided in the four locus scheme.

99

100

101

102 These more informative loci are not all internal fragments of housekeeping genes, as
103 frequently used in traditional MLST schemes. Two loci, *argE* and *gbpB*, are found within the coding
104 regions of genes (encoding acetylornithine deacetylase and a putative secreted antigen
105 respectively). The *gbpB* product is related to proteins which appear to be essential virulence factors
106 for other cocci (Teng et al., 2003; Stipp et al., 2008). *gbpB* in *M. plutonius* includes a VNTR (Variable
107 Number Tandem Repeat), which appears to be in an unstructured region of the protein. *argE* may be
108 important (Harris and Singer, 1998), though perhaps not essential (Kobayashi et al., 2003) for
109 arginine metabolism in *M. plutonius*. The remaining two loci, *galK* and *purR*, include intergenic
110 regions, and therefore cannot be considered traditional MLST loci. The *galK* locus spans a region
111 between two galactokinase fragments separated by a region of stop codons, and the *purR* locus
112 begins upstream of the 5' end of the *purR* (purine operon repressor) coding sequence. Whilst it is
113 therefore possible that *argE* and *gbpB* might have been located by a continued process of trial and
114 error of testing housekeeping genes, the *galK* and *purR* loci could only have been located by
115 comparing genomes for highly polymorphic regions.

116 Typing UK and International Isolates

117 The four locus typing scheme was tested on a further 41 isolates, sourced from the UK and
118 internationally. The STs identified so far reveal a total of 26 alleles over four loci (table 1). These
119 sequence data have been submitted to the EMBL database under the accession numbers HF569117
120 to HF569142. A list of isolates typed is provided in Supplementary Information.

121

122 Table 1. *M. plutonius* MLST typing scheme. Sequence Types (ST) identified to date are shown, and
123 the four columns, *galK* (galactokinase), *argE* (acetylornithine deacetylase), *gbpB* (secreted antigen)
124 and *purR* (purine operon repressor), show the alleles present at each of these four loci in each ST.
125 Within a column each number represents a unique allele (that is, a unique DNA sequence) at that
126 locus. Each ST has a unique allelic profile across all loci. The number of isolates found so far in each
127 ST is stated in the final column.

128

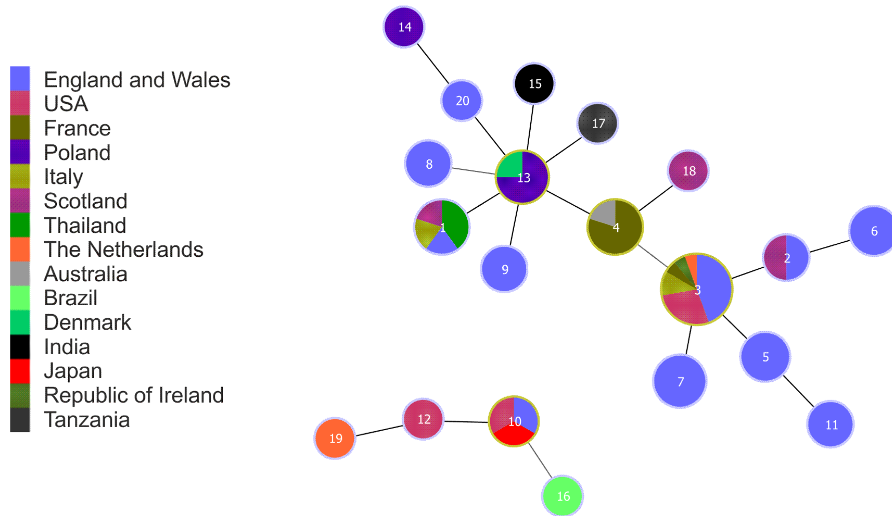
ST	*galK*	*argE*	*gbpB*	*purR*	Number of isolates
1	1	1	1	1	5
2	3	2	2	2	2
3	3	2	2	4	17
4	1	1	2	4	5
5	3	2	4	4	4
6	2	2	2	2	4
7	3	2	5	4	4
8	1	1	7	3	2
9	1	1	6	4	2
10	4	4	3	4	3
11	3	3	4	4	2
12	4	4	3	5	1
13	1	1	1	4	4
14	5	1	8	4	1
15	1	5	1	4	1
16	6	4	9	4	1

17	3	1	1	4	1
18	1	1	2	1	1
19	4	4	10	5	1
20	1	1	8	4	1

129

130

131



132
133 Figure 2. goeBURST Tree of STs. Each circle represents a different ST, with lines linking closest
134 relatives. Black lines indicate a single allelic change between STs, light grey lines indicate differences
135 at two loci. Circles ringed with a yellow outline indicate putative founder genotypes. Colours within
136 circles show the proportion of isolates of a particular type that were found in the countries
137 indicated.

138

139

140

141 In total, 12 STs were found in the UK, and 11 in the rest of the world (Figure 2). The
142 Sørensen similarity index was used to compare differentiation in ST distribution between the UK and
143 the rest of the world, as it requires only presence/absence data. The international isolates were not
144 sampled in a systematic way and therefore measures of ST abundance are not meaningfully
145 comparable with the UK. The Sørensen Index between the UK and the rest of the world is 0.2609,
146 which, on a 0-1 scale of similarity, indicates a low level of similarity between *M. plutonius*
147 populations. The measure β_{sim} (Lennon et al., 2001) takes into account the percentage of unshared
148 variants found in each area and performs well at reflecting gain and loss of species (Koleff et al.,
149 2003). The β_{sim} index for these samples is 0.7273, a high value of dissimilarity, and demonstrates that
150 the differentiation among communities is likely not due sampling efforts.

151 Isolates identical or similar by MLST to the Japanese atypical genome have been found in
152 Brazil, the UK, the USA and the Netherlands, suggesting that they are in fact widely distributed.
153 Interestingly, the Brazilian isolate (NCDO 2440) (Bailey, 1984) described in the 1990s (Allen and Ball,
154 1993) showed similar, but not identical, culture requirements to the Japanese atypical isolates (Arai
155 et al., 2012). Our MLST profile of this Brazilian isolate, ST16, places it in the atypical group, as it is
156 similar but not identical to that of the Japanese atypical isolate DAT 561 (ST10).

157 Local epidemiology

158 We have identified and sequence typed isolates from four outbreaks for which we have
159 detailed epidemiological data, in order to investigate the utility of this MLST scheme for confirming
160 links between infections.

161 In Case 1, a beekeeper with disease was known to have sold honey bees to a beekeeper 54
162 km away, and diseased material was sampled from both the seller (6th September 2010) and buyer
163 (29th June 2010). A 10 km grid square around the buyer had been free of EFB for the previous ten

164 years. Bacteria from both outbreaks were found to be ST9, an ST which was not found anywhere
165 else in the UK.

166 In Case 2, a different beekeeper with disease sold bees to another beekeeper 84 km away.
167 Bacteria from both seller (26th May 2010) and buyer (28th May 2010) outbreaks were ST11. Again,
168 this ST has been found nowhere else in the UK. The cases of ST9 and ST11 indicate that, when a rare
169 ST is identified, such information can be used as evidence to support hypothesised routes of
170 transmission.

171 In Case 3, diseased material was sampled from a shared apiary at three time points (27th
172 April 2010, 24th June 2010 and 21st June 2011). The bacteria recovered from all three outbreaks were
173 found to be ST6, a type only found at one other location in England, 3 km away from Case 3. The
174 case of ST6 is more complicated than the examples of the sale of bees. ST6 was seen to persist in the
175 apiary monitored in this study, despite treatment success at the colony level. There are two possible
176 explanations for this. Latent *M. plutonius* could be present in untreated asymptomatic colonies
177 (Budge et al., 2010). Such colonies would then have become symptomatic after the initial visit to
178 control the disease. However, the close proximity of the other ST6 apiary (3 km) indicates that this
179 could be an example of local, natural transmission events between the two sites. Here the typing
180 scheme has allowed us to generate data in support of hypotheses for infection of colonies that
181 would have been unsubstantiated without the genetic data.

182 Case 4 was comprised of a single apiary where all bees and equipment were imported into
183 the UK from Poland, six weeks prior to disease being observed in 40% of colonies. EFB had not been
184 found within 10 km of the apiary location in the previous 18 years. The bacteria recovered from this
185 outbreak were ST14, which was not isolated from any other UK samples. It is possible that this
186 instance depicts the import of novel STs into the UK. Three samples of *M. plutonius* were obtained
187 from Poland to try to confirm this, but all belonged to ST13. It is therefore clear that more extensive
188 sampling is required in the UK, and also Poland, before a firmer inference can be made.

189 The MLST scheme presented here has therefore shown that *M. plutonius* is not

190 genetically homogenous across its geographical range, and is able to discriminate multiple types

191 within groups of typical and atypical *M. plutonius*. Furthermore, the scheme's utility for making

192 epidemiological inferences has been demonstrated.

193 Acknowledgements

194 This project is funded by A BBSRC CASE studentship in partnership with Bee Disease Insurance and

195 the National Bee Unit. Assembly of contigs from 454 read data was performed by Rachel Glover. GB

196 was funded jointly by a grant from BBSRC, Defra, NERC, the Scottish Government and the Wellcome

197 Trust, under the Insect Pollinator Initiative (Grant BB/I000801/1). The authors gratefully

198 acknowledge the co-operation of all collaborators who provided isolates to the National Bee Unit

199 culture collection for this study.

200 References

201 Ahmed,N., Devi,S.M., de los Á Valverde,M., Viyachari,P., Machang'u,R.S., Ellis,W.A., and
202 Hartskeerl,R.A. (2006) Multilocus sequence typing method for identification and genotypic
203 classification of pathogenic *Leptospira* species. *Ann Clin Microbiol Antimicrob* **5**: 28

204 Allen,M.F. and Ball,B.V. (1993) The cultural characteristics and serological relationships of isolates of
205 *Melissococcus pluton*. *J Api Res* **32**: 80–88.

206 Arai,R., Tominaga,K., Wu,M., Okura,M., Ito,K., Okamura,N., et al. (2012) Diversity of *Melissococcus*
207 *plutonius* from Honeybee Larvae in Japan and Experimental Reproduction of European
208 Foulbrood with Cultured Atypical Isolates. *PloS one* **7**: e33708.

209 Bailey,L. (1984) A strain of *Melissococcus pluton* cultivable on chemically defined media. *FEMS*
210 *Microbiol Lett* **25**: 139–141.

211 Bailey,L. (1981) Honey Bee Pathology 1st ed. Academic Press, London.

212 Bailey,L. (1957) The isolation and cultural characteristics of *Streptococcus pluton* and further
213 Observations on *Bacterium eurydice*. *J Gen Microbiol* **17**: 39–48.

214 Budge,G.E., Barrett,B., Jones,B., Pietravalle,S., Marris,G., Chantawannakul,P., et al. (2010) The
215 occurrence of *Melissococcus plutonius* in healthy colonies of *Apis mellifera* and the efficacy of
216 European foulbrood control measures. *J Invertebr Pathol* **105**: 164–170.

217 Darling,A.E., Mau,B., and Perna,N.T. (2010) progressiveMauve: multiple genome alignment with
218 gene gain, loss and rearrangement. *PloS one* **5**: e11147.

219 Djordjevic,S.P., Smith,L.A., Forbes,W.A., and Hornitzky,M.A. (1999) Geographically diverse Australian
220 isolates of *Melissococcus pluton* exhibit minimal genotypic diversity by restriction
221 endonuclease analysis. *FEMS Microbiol Lett* **173**: 311–318.

222 Forsgren,E., Lundhagen,A.C., Imdorf,A., and Fries,I. (2005) Distribution of *Melissococcus plutonius* in
223 honeybee colonies with and without symptoms of European foulbrood. *Microbial Ecol* **50**: 369–
224 374.

225 Francisco,A.P., Bugalho,M., Ramirez,M., and Carrico,J.A. (2009) Global optimal eBURST analysis of
226 multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**: 152.

227 Gupta,S. and Maiden,M.C. (2001) Exploring the evolution of diversity in pathogen populations.
228 *Trends Microbiol* **9**: 181–5.

229 Harris,B.Z. and Singer,M. (1998) Identification and Characterization of the *Myxococcus xanthus* argE
230 Gene. *J Bacteriol* **180**: 6412–6414.

231 Homan,W.L., Tribe,D., Poznanski,S., Li,M., Hogg,G., Spalburg,E., et al. (2002) Multilocus Sequence
232 Typing Scheme for *Enterococcus faecium*. *J Clin Microbiol* **40**: 1963–1971.

233 Killgore,G., Thompson,A., Johnson,S., Brazier,J., Kuijper,E., Pepin,J., et al. (2008) Comparison of
234 Seven Techniques for Typing International Epidemic Strains of *Clostridium difficile*: Restriction
235 Endonuclease Analysis, Pulsed-Field Gel Electrophoresis, PCR-Ribotyping, Multilocus Sequence
236 Typing, Multilocus Variable-Number Tandem-Repeat Analysis, Amplified Fragment Length
237 Polymorphism, and Surface Layer Protein A Gene Sequence Typing. *J Clin Microbiol* **46**: 431–
238 437.

239 Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., et al. (2003) Essential
240 *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100**: 4678–83.

241 Koleff,P., Gaston,K.J., and Lennon,J.J. (2003) Measuring Beta Diversity for Presence–Absence Data. *J*
242 *Anim Ecol* **72**: 367–382.

243 Lennon,J.J., Koleff,P., Greenwood,J.J.D., and Gaston,K.J. (2001) The geographical structure of British
244 bird distributions: diversity, spatial turnover and scale. *J Anim Ecol* **70**: 966–979.

245 Maiden,M., Bygraves,J., Feil,E., Morelli,G., Russell,J., Urwin,R., et al. (1998) Multilocus sequence
246 typing: a portable approach to the identification of clones within populations of pathogenic
247 microorganisms. *Proc Natl Acad Sci USA* **95**: 3140–3145.

248 NBU (2012a) BeeBase - Inspection Reports. [WWW document]. URL
249 https://secure.fera.defra.gov.uk/beebase/public/BeeDiseases/apiaryReport.cfm

250 NBU (2012b) BeeBase - Trends of AFB and EFB. [WWW document]. URL
251 https://secure.fera.defra.gov.uk/beebase/public/BeeDiseases/trendDiseaseChart.cfm?id=10

252 Roetschi,A., Berthoud,H., Kuhn,R., and Imdorf,A. (2008) Infection rate based on quantitative real-
253 time PCR of *Melissococcus plutonius*, the causal agent of European foulbrood, in honeybee
254 colonies before and after apiary sanitation. *Apidologie* **39**: 362–371.

255 Ruiz-Garbajosa,P., Bonten,M.J.M., Robinson,D.A., Top,J., Nallapareddy,S.R., Torres,C., et al. (2006)
256 Multilocus Sequence Typing Scheme for *Enterococcus faecalis* Reveals Hospital-Adapted
257 Genetic Complexes in a Background of High Rates of Recombination. *J Clin Microbiol* **44**: 2220–
258 2228.

259 Stipp,R.N., Gonçalves,R.B., Höfling,J.F., Smith,D.J., and Mattos-Graner,R.O. (2008) Transcriptional
260 analysis of gtfB, gtfC, and gbpB and their putative response regulators in several isolates of
261 *Streptococcus mutans*. *Oral Microbiol Immunol* **23**: 466–473.

262 Teng,F., Kawalec,M., Weinstock,G.M., Hryniewicz,W., and Murray,B.E. (2003) An *Enterococcus*
263 *faecium* Secreted Antigen, SagA, Exhibits Broad-Spectrum Binding to Extracellular Matrix
264 Proteins and Appears Essential for *E. faecium* Growth. *Infect Immun* **71**: 5033–5041.

265 Urwin,R. and Maiden,M. (2003) Multi-locus sequence typing: a tool for global epidemiology. *Trends*
266 *Microbiol* **11**: 479–487.

267 White,G.F. (1912) The Cause of European Foul Brood. *Circ U S Dep Agric* **157**: 1–22.

268 Wilkins,S., Brown,M.A., and Cuthbertson,A.G.S. (2007) The incidence of honey bee pests and
269 diseases in England and Wales. *Pest Manag Sci* **63**: 1062–1068.

270

271

272 Supplementary Information

273 List of Isolates

Fera ID	Alternative ID	Country of Origin	Donated By	ST
6404*	LMG 20360	England and Wales		1
7087*		England and Wales		3
7102*		England and Wales		2
7148*		Thailand	1	1
7149		Thailand	1	1
7154*	NCFB 2442	Australia		4
7363		France		4
7365		France		4

7366	France	4
7369	France	4
7483*	England and Wales	7
7509*	England and Wales	3
7511*	England and Wales	7
7512†	England and Wales	6
7515*	England and Wales	7
7516*	England and Wales	3
7517*	England and Wales	3
7521*	Scotland	2
7523*	England and Wales	3
7524*	England and Wales	3
7526*†	England and Wales	6
7531*	England and Wales	7
7533†	England and Wales	6
7534*†	England and Wales	8
7540*†	England and Wales	9
7595†	England and Wales	6
7596*	England and Wales	10
7599*	England and Wales	3
7604	England and Wales	8
7605†	England and Wales	11
7606	England and Wales	5
7609*	England and Wales	5
7611†	England and Wales	11
7612	England and Wales	5
7613†	England and Wales	5

7641		England and Wales		9
8061		England and Wales		20
8214		USA	2	3
8217		USA	2	10
8220		USA	2	3
8222		USA	2	3
8224		USA	2	3
8396		Denmark	3	13
8397		Poland (sampled in UK)		14
8401		Republic of Ireland	4	3
8469	NCFB 2440	Brazil		16
8470	NCFB 2439	India		15
8472	NCFB 2441	Tanzania		17
8473		Scotland	5	18
8475		Scotland	5	1
8498		The Netherlands	6	3
8500		France		3
8513		The Netherlands	6	19
8516		Poland	7	13
8517		Poland	7	13
	EFB C+70818	Italy	8	1
	EFB 1185/1	Italy	8	3
	EFB 1185/2	Italy	8	3
	US21	USA	2	3
	US18	USA	2	12
	P4	Poland	7	13
	DAT 561	Japan	‡	10

274

275 Table showing isolates used in this study. Samples without a Fera ID were not cultured, and PCR was
276 performed on a whole larval DNA extract, or on a culture extract supplied by the donor. Isolates
277 highlighted with a * are those that form the initial 21 isolates on which all loci were screened.
278 Isolates highlighted with a † are those that were used for genomic comparisons.

279 Donors;

280 1) Dr Panuwan Chantawannakul, Department of Biology, Faculty of Science, Chiang Mai
281 University, Chiang Mai, Thailand

282 2) Prof Jeff Pettis and Mr I Barton Smith, Bee Research Laboratory, USDA, Beltsville, MD, USA

283 3) Dr Eva Forsgren, Department of Ecology, Swedish University of Agricultural Sciences,
284 Uppsala, Sweden

285 4) Mr Eoghan Mac Giolla Coda, Galtee Bee Breeding Group, Republic of Ireland

286 5) Dr Fiona Highet and Ms Mairi Carnegie, Science and Advice for Scottish Agriculture (SASA),
287 Edinburgh, UK

288 6) Dr Sjef van der Steen and Dr Bram Cornelissen, Wageningen University, Netherlands

289 7) Dr Krystyna Pohorecka and Mr Andrzej Bober, National Veterinary Research Institute,
290 Pulawy, Poland

291 8) Dr Anna Granato, Istituto Zooprofilattico Sperimentale delle Venezie, viale dell'Università,
292 Legnaro, Italy

293

294 ‡ Allele sequences from strain DAT 561 are taken from NC_016938.1. Note; at positions 36-41 in all
295 *purR* alleles except that from NC_016938.1, 6 thymine bases are present. In NC_016938.1 there is a
296 seventh thymine inserted, but this is believed to be an artefact of sequence assembly, as 454
297 technology can have difficulty resolving long tandem sequences of thymines. This assumption is
298 supported by the fact that the type strain sequence published by the same group (NC_015516) also
299 has 7 thymine bases present, whereas our Sanger sequencing of the type strain allele shows 6
300 thymine bases. We have therefore not treated this extra thymine in DAT 561 as indicating a novel
301 allele.

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317 Experimental Procedures

318 Bacterial growth and DNA extraction

319 Isolates of *M. plutonius* from the Fera culture collection were plated out on to M110 agar

320 and grown under anaerobic conditions at 35°C ± 2°C. To isolate bacteria from individual infected

321 larvae, larvae were mixed with 0.5 ml PBS and vortexed, with a 10µl loop used to streak the resulting

322 suspension on M110 plates and incubated as before. After growth, individual colonies were spread

323 on to fresh M110 plates, and left to grow for 7-14 days (Forsgren et al., 2013)

324 DNA was extracted from *M. plutonius* colonies using the Promega Wizard ® Magnetic DNA

325 Purification System for Food. In some cases, diseased larvae did not yield viable bacteria, so a

326 suspension of larva in PBS underwent DNA extraction by the same method.

327

328 Identification of loci

329 Loci were tested for the same genes as the MLST schemes for *Enterococcus faecalis* (Ruiz-
330 Garbajosa et al., 2006) and *Enterococcus faecium* (Homan et al., 2002), and for genes that had
331 predicted membrane-spanning domains. Subsequently loci were identified based on Mauve (Darling
332 et al., 2010) alignments between contigs of *M. plutonius* genome sequence, from 9 isolates collected
333 in England in 2010 (contigs having previously been produced using newbler (Roche) with reads from
334 a Roche 454 pyrosequencer (data not shown)). The Export SNPs function of Mauve was used to
335 locate polymorphic sites. These were verified by manual BLASTn alignments between contigs from
336 all isolates. Loci were chosen that showed three or more different sequences between 10 isolates
337 over a 300bp region. Primers were designed to the *M. plutonius* type strain sequence, using primer3
338 (Rozen and Skaletsky, 2000), and were synthesised by MWG Eurofins.

339 Testing of locus variability

340 PCR primers were first used to amplify fragments from DNA from a test set of *M. plutonius*
341 isolates, including culture-collection isolates and those isolated from the field in the UK. PCR
342 products were purified using the Qiagen Qiaquick PCR purification kit, and sequenced in one
343 direction by Sanger sequencing at MWG Eurofins. Sequencing was performed on an ABI 3730xl 96-
344 capillary DNA Analyzer.

345 Loci based on previous MLST schemes and the presence of membrane-spanning domains did
346 not show sufficient variation to merit inclusion in a scheme. Loci identified from 454 data, *galK*,
347 *argE*, *GbpB*, and *PurR*, did reveal higher levels of variation, and were able to distinguish between
348 multiple STs. These loci were then used as the final version of the modified MLST scheme, and were
349 used to sequence type all further isolates.

350 Primer Sequences and PCR Conditions

351 PCR conditions;

352 94°C for 2 minutes; 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 minute;

353 final elongation at 72°C for 1 minute.

354

355 Primer sequences;

356 galK L (TTT CCA GCA GCA ATT ACA A) and galK R (GGG TAG GGA TTT TTG AAG AG)

357 argE L (GGT GGG ACA TTT AGA CGT AG) and argE R (AAA TTA AGA CCC AAC CCT TC)

358 GbpB L (AGC AGC TAA ACA GAA TGA GC) and GbpB R (GCC AAC GTC TAA CAG ATA CC)

359 PurR L (ACC ACC AAG TGC CAG TAT TA) and PurR R (CGA TTT TGT TCT GAT AAC CTG)

360

361 International Isolates

362 After initial validation on a test set of culture collection and UK-collected isolates, a range of

363 globally sourced samples were tested with the MLST primers. Samples were tested from Australia,

364 Brazil, Denmark, France, India, Italy, the Netherlands, Poland, the Republic of Ireland, Tanzania,

365 Thailand, and the United States of America, and sequences for the MLST loci were downloaded from

366 the Genbank genome sequence of a sample of atypical *M. plutonius* from Japan (NC_016938.1).

367 Case Studies of Suspected Honey bee Movements

368 To demonstrate the effectiveness of the MLST scheme for making epidemiological

369 inferences, several anonymised cases of EFB linked to known or suspected honey bee movements or

370 disease recurrence were identified by discussion with National Bee Unit field staff (G. Budge, *Pers.*

371 *comm.*). *M. plutonius* cultures were grown from infected larvae taken from these outbreaks, and

372 typed. All case studies occurred in 2010.

373 Data Analysis

374 Potential patterns of evolutionary descent between STs were calculated using the goeBURST
375 algorithm in the PHYLOViZ program (Francisco et al., 2012) a modification of the earlier eBURST
376 algorithm (Feil et al., 2004). In figure 1 the two goeBURST plots were merged in corelDRAW, as was
377 the colour key in figure 2. Measures of differentiation between STs present in the UK and the rest of
378 the world were calculated in R (R Core Team, 2012) using the vegan package (Okansen et al., 2012).

379 References

380 Darling,A.E., Mau,B., and Perna,N.T. (2010) progressiveMauve: multiple genome alignment with
381 gene gain, loss and rearrangement. *PloS one* **5**: e11147.

382 Feil,E.J., Li,B.C., Aanensen,D.M., Hanage,W.P., and Spratt,B.G. (2004) eBURST: Inferring Patterns of
383 Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence
384 Typing Data. *J Bacteriol* **186**: 1518–1530.

385 Forsgren,E., Budge,G.E., Charrière,J.-D., and Hornitzsky,M.A.Z. (2013) Standard methods for
386 European foulbrood research. *J Api Res* **52** (1)

387 Francisco,A.P., Vaz,C., Monteiro,P.T., Melo-Cristino,J., Ramirez,M., and Carriço,J.A. (2012) PHYLOViZ:
388 phylogenetic inference and data visualization for sequence based typing methods. *BMC*
389 *Bioinformatics* **13**: 87.

390 Homan,W.L., Tribe,D., Poznanski,S., Li,M., Hogg,G., Spalburg,E., et al. (2002) Multilocus Sequence
391 Typing Scheme for Enterococcus faecium. *J Clin Microbiol* **40**: 1963–1971.

392 Okansen,J., Guillaume Blanchet,F., Kindt,R., Legendre,P., Minchin,P.R., O'Hara,R.B., et al. (2012)
393 vegan: Community Ecology Package. [WWW document] URL http://CRAN.R-
394 project.org/package=vegan

395

396 R Core Team (2012) R: A Language and Environment for Statistical Computing [WWW document]
397 URL http://www.R-project.org/.

398 Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist
399 programmers. *Methods Mol Biol*.

400 Ruiz-Garbajosa,P., Bonten,M.J.M., Robinson,D.A., Top,J., Nallapareddy,S.R., Torres,C., et al. (2006)
401 Multilocus Sequence Typing Scheme for Enterococcus faecalis Reveals Hospital-Adapted
402 Genetic Complexes in a Background of High Rates of Recombination. *J Clin Microbiol* **44**: 2220–
403 2228.

404

405