Heading title: Diversity of *Xanthomonas translucens* in Iran

Molecular Typing Reveals High Genetic Diversity of *Xanthomonas translucens* Infecting Small-Grain Cereals in Iran

Moein Khojasteh[1]; S. Mohsen Taghavi[1]; Pejman Khodaygan[2]; Habiballah Hamzehzarghani[1]; Gongyou Chen[3]; Claude Bragard[4]; Ralf Koebnik[5]; Ebrahim Osdaghi[1]*

[1]Department of Plant Protection, College of Agriculture, Shiraz University, Shiraz 71441-65186, Iran

[2]Department of Plant Protection, Faculty of Agriculture, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

[3]State Key Laboratory of Microbial Metabolism, School of Life Science & Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China

[4]Laboratory of Phytopathology-Applied Microbiology, Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium

[5]IRD, Cirad, Univ. Montpellier, Interactions Plantes Microorganismes Environnement (IPME), Montpellier, France

*For correspondence: eosdaghi@shirazu.ac.ir

Abstract

This study provides a phylogeographic insight into the population diversity of *Xanthomonas translucens* strains causing bacterial leaf streak disease of small-grain cereals in Iran. Among the 65 bacterial strains isolated from wheat, barley, and gramineous weeds in eight Iranian provinces, using multilocus sequence analyses and typing (MLSA/MLST) of four housekeeping genes (*i.e. dnaK*, *fyuA*, *gyrB*, and *rpoD*), 57 strains were identified as *X. translucens* pv. *undulosa*, while eight strains were identified as *X. translucens* pv. *translucens*. Although the pathogenicity patterns on oat and ryegrass weed species were varied among the strains, all *X. translucens* pv. *undulosa* strains were pathogenic on barley, Harding's grass, rye (except for XtKm35) and wheat, and all *X. translucens* pv. *translucens* strains were pathogenic on barley and Harding's grass, while none of the latter group was pathogenic on rye and wheat (except for XtKm18). MLST using the 65 strains isolated in Iran as well as the sequences of the four genes from 112 strains of worldwide origin retrieved from the GenBank database revealed higher genetic diversity (*i.e.* haplotype frequency, haplotype diversity, and percentage of polymorphic sites) among the Iranian population of *X. translucens* in comparison to the North American strains of the pathogen. High genetic diversity of the BLS pathogen in Iran was in congruence with the fact that the Iranian Plateau is considered the center of origin of cultivated wheat. However, further studies using a larger collection of strains are warranted to precisely elucidate the global population diversity and center of origin of the pathogen.

41 **Importance**

42 Bacterial leaf streak (BLS) of small-grain cereals (*i.e.* wheat and barley) is one of the
43 economically important diseases of gramineous crops worldwide. The disease occurs in many
44 countries across the globe with a particular importance in regions characterized by high
45 precipitations. Two genetically distinct xanthomonads – namely *X. translucens* pv. *undulosa*
46 and *X. translucens* pv. *translucens* – have been reported to cause BLS disease on small-grain
47 cereals. As a seed-borne pathogen, the causal agents are included in the A2 list of quarantine
48 pathogens by the European and Mediterranean Plant Protection Organization (EPPO). Despite
49 its global distribution and high economic importance, population structure, genetic diversity,
50 and phylogeography of *X. translucens* remain undetermined. This study, using MLSA/MLST
51 provides a global-scale phylogeography of small-grain cereals infecting *X. translucens*. Based
52 on the diversity parameters, neutrality indices, and population structure we indicate higher
53 genetic diversity of BLS pathogen in Iran − which is geographically close to the center of origin
54 of common wheat – than that has so far been observed in the other areas of the world
55 including North America. Results obtained in this study provide a novel insight into the
56 genetic diversity and population structure of the BLS pathogen of small grain cereals in a
57 global scale.

58 **Key words:** Bacterial leaf streak, *Hordeum vulgare*, Iranian Plateau, MLSA/MLST,
59 Phylogeography, *Triticum aestivum*

60

61

62 **Introduction**

63 Among the biotic constraints of small-grain cereals in the *Poaceae* family, bacterial leaf streak
64 (BLS) caused by different pathovars of *Xanthomonas translucens* (*i.e. X. translucens* pv.
65 *translucens* and *X. translucens* pv. *undulosa*) is considered one of the most important seed-
66 borne pathogens around the globe (1−2). The disease was first described on barley (*Hordeum*
67 *vulgare*) in the USA in 1917, and currently the causal agent is widespread through all six
68 continents. *Xanthomonas translucens* is included in the A2 list of quarantine pathogens by the
69 European and Mediterranean Plant Protection Organization (3). The pathogen is seed-borne
70 and infected seeds are the main source of inoculum for long-distance distribution.
71 Furthermore, a humid and warm environment coupled with frequent overhead irrigation
72 maximizes bacterial growth and plant to plant spread in the field, leading to 10-40% yield
73 losses (2). When the symptoms are found on wheat glumes (a leaf-like structure attached to
74 the flowers of Gramineae) the disease is called “black chaff”, and characterized by black,
75 longitudinal, and parallel stripes (1). Yield loss due to the BLS is proposed to be attributed to
76 disease severity on the wheat spikes and flag leaf (the uppermost leaf of the plant). In severe
77 cases of disease incidence, 5-10% of the wheat spikes may be sterile due to the BLS infection.
78 It has also been noted that 50 and 100% disease severity on the flag leaf will be resulted in an
79 8-13% and 13-34% kernel weight losses, respectively (2).

80 Until recently, discrimination of *X. translucens* pathovars has mostly been relied on the host
81 range and symptom profiles of the pathogens on a set of gramineous species (4−6). For
82 instance, it has been noted that *X. translucens* pv. *translucens* − including the strains
83 previously designated as *X. translucens* pv. *hordei* − is limited to barley, while *X. translucens*

84 pv. *undulosa* displays a broader host range infecting wheat, barley, oat, rye, bromes and
85 triticale (2, 5). Pathovar designation within *X. translucens* is complicated not only because of
86 the overlapping host ranges (5, 7), but also because of similar genotypic features of the
87 pathotype strains (3, 6). Conventional bacteriological typing techniques (*e.g.* rep-PCR, AFLP,
88 and protein/fatty acid profiling) can only partly solve the taxonomic issues within the *X.*
89 *translucens* pathovars (4–5, 8–9). However, it has recently been shown that multilocus
90 sequence analyses (MLSA) and typing (MLST) using partial sequences of four housekeeping
91 genes (*i.e. dnaK*, *fyuA*, *gyrB*, and *rpoD*) could discriminate *X. translucens* strains isolated from
92 gramineous crops, and the resulting scheme was in congruence with their experimental host
93 range, host of isolation, as well as draft genome sequence-based phylogeny of the strains (3,
94 6).

95 Although Curland et al. (6) pinpointed − so far − the best method for evaluating genetic
96 diversity of *X. translucens*, the exploration was limited to North American strains, thus leaving
97 the "Old World" population of the pathogen uninvestigated. Considering the center of origin
98 of cultivated wheat which was determined to be in the Karacadağ Mountains in southeast
99 Turkey (10–11) a large-scale phylogenetic analysis is warranted to shed light on the
100 population structure and diversity parameters of *X. translucens* in the Old World, and to
101 provide a global-scale phylogeography of the species. While field symptoms resembling BLS
102 infection have occasionally been observed for a long time on barley and wheat in Iran, the
103 presence of *X. translucens* was scientifically first described in Kerman province (southern Iran)
104 in 1983 (12). However, information on the geographic distribution and phylogenetic position
105 of the causal agent(s) is mostly lacking in the country. Given the fact that wheat is a staple

106 food in Iran with highest economic importance among the agricultural commodities, precise
107 characterization and identification of the BLS pathogen will aim to develop sustainable
108 methods to manage the disease. Indeed, by 2016, wheat cultivation occupied over 50% of the
109 total area dedicated to annual crops in Iran, and the areas cultivated with wheat accounted
110 for more than 5,928,000 ha with an annual production of 14.6 million tonnes (13). Hence, a
111 multi-phasic characterization and comparative phylogenetic analysis is warranted to decipher
112 the population structure of the BLS pathogen in Iran, which at the same time will contribute
113 to elucidate the phylogeography of cereal-infecting *X. translucens* strains at a global scale.

114 The objectives of the present study were to I) assess the host range of *X. translucens* strains
115 isolated from wheat, barley, rye, and ryegrass (*Lolium* spp.) in Iran, and II) determine the
116 phylogenetic position of the strains using the four-gene MLSA/MLST scheme (*i.e. dnaK*, *fyuA*,
117 *gyrB*, and *rpoD*). Furthermore, a global-scale phylogeographic analysis was conducted using
118 the strains isolated in Iran and the publicly available sequences of 112 *X. translucens* strains of
119 worldwide origin (6, 14). Haplotype diversity and sequence variation statistics showed that
120 the strains isolated in Iran are possessing higher genetic diversity in comparison to the New
121 World strains of the pathogen.

122 **Results**

123 **Bacterial Strains and Host Range Assays**

124 Sixty-five Gram-negative xanthomonad-like bacterial strains were isolated from wheat, barley
125 and gramineous weeds in cereal-growing areas in Iran from 2008 to 2017 (Table 1).
126 Abundance of the surveyed fields in each area/year depended on the BLS incidence and

127 severity reports received from the local crop sanitary inspectors. The 65 strains were isolated
128 in eight provinces thorough southern (*i.e.* Bushehr, Fars, and Kerman), western (*i.e.* Kurdistan,
129 Hamadan, Lorestan, and Kermanshah) and northwestern (*i.e.* Zanjan) Iran (Figure S1). All the
130 strains formed pale-yellow colonies on YDC medium and elicited a hypersensitivity reaction
131 (HR) on tobacco leaves. Furthermore, the specific primer pair T1/T2 (7) amplified a 139 bp
132 DNA fragment of *tRNA-Ala* locus in all 65 strains that was useful in identifying *X. translucens*
133 species. Consequently, all the strains were subjected to the pathogenicity tests and host rang
134 assays.

135 In the host range assays, all strains were shown to be pathogenic on their host of isolation
136 except for the strain XtKm15, which was isolated from ryegrass but did not induce symptoms
137 on this species. On the other hand, all but seven strains (*i.e.* XtKm4, XtKm7, XtKm8, XtKm9,
138 XtKm11, XtKm33, and XtKm34) were pathogenic on wheat, although the symptoms produced
139 by XtKm18 on wheat plants were less severe than those produced by the others. All the
140 strains that were non-pathogenic on wheat plants had been isolated from barley. Although all
141 the strains were pathogenic on barley plants, the barley strain XtKm9 as well as the type
142 strain of *X. translucens* pv. *translucens* (DSM 18974, used as positive control) were
143 pathogenic only on barley plants. All the strains isolated from wheat were pathogenic on
144 wheat and barley, while only half of the barley strains (7 out of 14) were pathogenic on wheat
145 (Table 1; Figure 1). In general, the host range of the wheat strains was broader than that of
146 the barley strains. Twenty wheat strains, four barley strains (*i.e.* XtHn9, XtHn11, XtLr7, and
147 XtLr9), as well as the rye strain XtKm3 were pathogenic on all six tested small-grain cereals

148 (Figure 1). The few remaining strains were non-pathogenic on at least one of the tested plant
149 species in the greenhouse conditions.

150 **Phylogenetic Analyses**

151 Multilocus sequence analysis of four housekeeping genes *(i.e. dnaK, fyuA, gyrB,* and *rpoD*)
152 was conducted on all 65 strains used in this study. BLAST searches using the sequences of
153 each of the individual genes revealed that the strains isolated in Iran had 99-100% sequence
154 identity with the type/reference strains of *X. translucens i.e. X. translucens* pv. *undulosa* (LMG
155 892^T), *X. translucens* pv. *secalis* (LMG 883^T) and *X. translucens* pv. *translucens* (ICMP 5752^T).
156 Phylogenetic tree constructed using the dataset of concatenated sequences of *dnaK*, *fyuA*,
157 *gyrB* and *rpoD* genes confirmed the results obtained from the BLAST analyses. All the strains
158 isolated in Iran clustered in a monophyletic clade within *X. translucens*, although they were
159 scattered through different pathovars and phylogroups of the species (Figure 1; Figure S2).
160 Similar results were obtained when the sequences of individual housekeeping genes were
161 subjected to the phylogenetic analysis (data not shown).

162 Eighty-seven percent of *X. translucens* strains isolated in Iran (57 out of 65 strains) were
163 clustered with the pathotype strains of *X. translucens* pv. *undulosa* and *X. translucens* pv.
164 *secalis*. These strains were divided into two main sub-clusters (*i.e.* sub-cluster I including 23
165 strains and sub-cluster II including 18 strains) as well as 16 un-clustered strains (Figure 1).
166 Among the aforementioned 57 strains, 49 strains were isolated from wheat, six strains were
167 isolated from barley and each of the two remaining strains were isolated from rye or from
168 ryegrass. According to the host of isolation, we refer to the wheat, barley and ryegrass strains

as *X. translucens* pv. *undulosa,* while the rye strain (XtKm3) is considered *X. translucens* pv. *secalis.* Regardless of their geographic origin (province/county) all 49 wheat strains as well as the two strains isolated from weed species (*i.e.* XtKm3 and XtKm15 isolated from rye and ryegrass, respectively) were identified as *X. translucens* pv. *undulosa*, while only 43% (6 out of 14) of the barley strains were clustered in this pathovar (Figure 1). The remaining eight barley strains were clustered among the strains of *X. translucens* pv. *translucens*, however; they were scattered through three sub-clusters. Interestingly, all the Iranian strains identified as *X. translucens* pv. *translucens* were isolated from barley in the Kerman province (Figure S3). Although the strains XtKm18, XtKm33, XtKm34, XtKm8 and XtKm7 were clustered in a monophyletic group, XtKm7 was differentiated from the remaining four strains and placed in a separate sub-cluster (Figure 1). The strain XtKm7 showed one and five nucleotide differences in the *gyrB* and *rpoD* genes sequences, respectively, compared to the sub-cluster consisting of XtKm8, XtKm18, XtKm33, and XtKm34 strains. Furthermore, the strains XtKm11, XtKm4, and XtKm9 which clustered together with the strains isolated from barley in the USA, had one, three and six nucleotide differences in the *dnaK*, *fyuA*, and *rpoD* gene sequences, respectively, with the subgroup containing XtKm8, XtKm18, XtKm33, and XtKm34 strains. Moreover, there were three nucleotides differences in the *fyuA* sequences and one nucleotide differences in *dnaK*, *gyrB*, and *rpoD* sequences between the strain XtKm7 and the three strains XtKm4, XtKm9, and XtKm11.

Phylogenetic position of the strains was in congruence not only with their host of isolation but also with the experimental host range of the strains. For instance, all the strains identified as *X. translucens* pv. *undulosa* were pathogenic on wheat, barley, Harding's grass (*Phalaris*

191 *aquatica*), and rye except for the wheat strain XtKm35, which was not pathogenic on rye.
192 Whereas all *X. translucens* pv. *translucens* strains were pathogenic on barley and Harding's
193 grass, and none of them was pathogenic on wheat and rye except for the strain XtKm18,
194 which induced water soaking symptoms on wheat plants. Pathogenicity patterns on ryegrass
195 and oat plants were variable among the strains regardless of their host of isolation and
196 phylogenetic position. Interestingly, all the 23 strains which clustered in the sub-cluster I of *X.*
197 *translucens* pv. *undulosa* were pathogenic on all the six evaluated plant species. However,
198 neither oat nor ryegrass plants were infected with the 18 strains in sub-cluster II. More
199 specifically, none of the strains which included XtFa2, XtFa3, XtHn2, XtKm10, XtKm22,
200 XtKm24, XtKm30, XtKr1, XtLr1, and XtZa1, and which were in the same multilocus haplotype
201 (MH), was pathogenic on oat.

202 **Genetic Diversity**

203 The strains isolated in Iran carried different allelic forms and sequence types (STs) and
204 corresponded to 26 MHs based on the concatenated sequences of four housekeeping genes
205 (Table S1). As for the individual genes in the Iranian strains, fifteen, three, five, and eleven STs
206 were detected for the *dnaK*, *fyuA*, *gyrB*, and *rpoD* genes, respectively (Figure S3), while only
207 eight, three, two, and eight STs were detected for these genes in 112 strains from other
208 countries (Table S1). Sequence variation statistics and diversity parameters among the strains
209 isolated in Iran were calculated using the DnaSP 5.10 software and compared with those of
210 112 *X. translucens* strains isolated in six different (non-Iran) countries (Table 2). As for the 58
211 *X. translucens* pv. *undulosa* strains isolated in Iran (*i.e.* 57 strains isolated in this study and the
212 reference strain ICMP 11055 isolated in Sothern Iran in 1983), there were eight, one, four,

213 and seven STs in the sequences of *dnaK*, *fyuA*, *gyrB*, and *rpoD* genes, respectively, while
214 seven, two, three, and four STs, respectively, were found in the *X. translucens* pv. *translucens*
215 strains (Figure 2). Haplotype frequency indices (HF: ratio of the "number of haplotypes" per
216 "number of strains") were 0.137, 0.017, 0.068, and 0.120 for *X. translucens* pv. *undulosa* and
217 0.875, 0.250, 0.375, and 0.500 for *X. translucens* pv. *translucens* strains in the *dnaK*, *fyuA*,
218 *gyrB*, and *rpoD* genes, respectively, indicating a higher genetic diversity among the *X.*
219 *translucens* pv. *translucens* strains in Iran. Furthermore, comparison of HF and haplotype
220 diversity (HD) indices between the strains isolated in Iran with those of 112 non-Iran strains
221 revealed a higher genetic diversity among the Iranian strains of *X. translucens* (Table 2).

222 Several strains isolated in different geographical areas in Iran shared the same MH with
223 strains isolated in other countries (Figure 2). For instance, strain XtKm3 isolated in Kerman
224 (Iran) as well as strains CIX36, CIX53, CIX54 isolated in Minnesota (USA) belonged to the same
225 MH, while strain XtKm9 shared the same MH with the barley strain UPB458 isolated in India.
226 Interestingly, MH2 has include 15 and 39 strains isolated in Iran and the USA, respectively, as
227 well as the pathotype strains of *X. translucens* pv. *undulosa* (LMG 892^T) and *X. translucens* pv.
228 *secalis* (LMG 883^T). On the other hand, several MHs were represented by only one Iranian
229 strain (*e.g.* XtKm4, XtKm7, XtKm8, XtKm11, XtKm18, XtKm33, and XtKm34 for *X. translucens*
230 pv. *translucens* and XtFa4, XtHn7, XtKr4, XtLr5, XtLr6, XtLr8, and XtZa4 for *X. translucens* pv.
231 *undulosa*), which indicates a higher allelic richness of the pathogen in Iran (Table S1; Figure 2).

232 Phylogenetic networks were generated using the Splits-decomposition method for all the
233 individual genes, as well as for the concatenated sequences data set. When 178 worldwide
234 strains were analyzed, in spite of minor reticulations in the splits-decomposition network, the

pairwise homoplasy index (PHI) test did not find statistically significant evidence for recombination neither in the individual genes nor in concatenated datasets (data not shown). Results of the PHI test were in congruence with those of the Recombination Detection Program (RDP), which did not find statistically significant evidence for recombination. On the other hand, none of the population neutrality indices (*i.e.* Tajima's *D*, Fu and Li's D^*, Fu and Li's F^*) was significant among the 66 strains isolated in Iran, thus indicating the lack of evidence for selection among the population and that the population is evolving per mutation-drift equilibrium (Table 2). On the other hand, the non-Iran *X. translucens* pv. *undulosa* strains showed significant departure from the mutation/drift equilibrium in *rpoD* gene sequence. The negative value of neutrality indices in *rpoD* gene sequence of non-Iran *X. translucens* pv. *undulosa* strains suggests a recent selective sweep or/and population expansion after a recent bottleneck, although further investigations using the sequences of additional housekeeping genes are needed to prove these observations.

Phylogeography of Strains of Worldwide Origin

Several lines of evidence of geographic structure in the ST distribution were detected in the TCS networks generated by PopART v. 1.7 program (Figure 3). Although a number of STs were shared among the strains of worldwide origin, country-specific STs were more prevalent in the *dnaK* and *rpoD* networks. For instance, in the *dnaK*, *gyrB*, and *rpoD* sequences, there were eleven, four, and seven STs consisting of Iranian strains; while only two and four STs in the *dnaK* and *rpoD* sequences, respectively, corresponded to USA strains (Figure 3). On the other hand, the strain DAR61454 isolated in Australia shared the same MH with the strains isolated in Iran, Canada, and the USA; while the strain UPB458 isolated in India shared the same MH

257 with the strains isolated in southern Iran (Figure 2, Table S1), raising the speculation that
258 strain UPB458 was originated from Iran. Furthermore, six strains isolated in Canada (*i.e.*
259 XT5523, XT5770, XT5791, XT8, LMG 883 and LMG 892) as well as the strains UPB513 and
260 UPB787 isolated in Mexico and Paraguay, respectively, shared the same MH with the strains
261 isolated in Midwestern USA (Figure S4). The patterns of MHs positioning in TCS network
262 (Figures 2, 3) were in congruence with the MrBayes bifurcating trees; however, reticulations
263 in the TCS networks of the *dnaK* and *rpoD* sequences as well as of the concatenated
264 sequences suggest a more complex evolutionary history of *X. translucens* than that can be
265 observed in the phylogenetic trees (Figure 1).

266 Using the eBURST analysis, 178 *X. translucens* strains were assigned to 46 MHs within three
267 groups and four singleton sequences. Group one consisted of 143 strains and 37 MHs (Figure
268 S5), while the groups two and three – both consisted of only *X. translucens* pv. *translucens*
269 strains − included 30 and two strains, as well as four and two MHs, respectively (data not
270 shown). Although the founder genotype of *X. translucens* pv. *undulosa* (MH2) and the
271 respected subgroups in group one were supported with 99-100% bootstrap values (Figure S5),
272 assignment of the founder genotype in groups two and three was not supported by high
273 bootstrap value (>20), hence they were not considered reliable founder genotypes. The
274 founder genotype MH2 included 56 strains (as shown in Figure 2) with 20, nine, and four
275 single-, double- and triple-locus variants, respectively. Two single-locus variants of MH2 (*i.e.*
276 MH10 and MH28) were identified as the subgroup founders, each with its own cluster of
277 linked single-locus variants (Figure S5). Among the 56 strains in the MH2, 15 strains were
278 isolated in Iran while 40 and one strains were isolated in the USA and Australia, respectively.

279 The results obtained with eBURST analyses were in congruence with those of POPArt
280 analyses, suggesting MH2 as the primary founder genotype for *X. translucens* pv. *undulosa*
281 (Figure 2; Figure S5).

282

283 **Discussion**

284 This study describes the host range and phylogenetic position of 65 *X. translucens* strains
285 isolated from gramineous plants in Iran among the worldwide population of the pathogen. All
286 the wheat strains isolated in Iran were identified as *X. translucens* pv. *undulosa* while the
287 strains isolated from barley were shown to belong to two pathovars (*i.e. X. translucens* pv.
288 *undulosa* and *X. translucens* pv. *translucens*). None but one (XtKm18) of the *X. translucens* pv.
289 *translucens* strains isolated in Iran was pathogenic on wheat, while all the strains were
290 pathogenic on barley and Harding's grass. Relatively higher number of wheat strains (49/65)
291 isolated in our surveys might be attributed to the differences in the cultivated areas by the
292 two crops, and proportionally higher number of wheat fields surveyed in comparison to the
293 barley fields. Similar results, in terms of multi-species etiology and variable host range were
294 observed in the bacterial agents infecting solanaceous vegetables (15) edible dry beans (16)
295 and sugar beets (17) in Iran.

296 As revealed by several diversity parameters (*i.e.* haplotype frequency, haplotype diversity,
297 and percentage of polymorphic sites), strains isolated in Iran possess significantly higher
298 genetic diversity than the strains isolated in the North American countries. Furthermore, a
299 global-scale phylogeographic analysis supports the hypothesis that Iranian strains of *X.*

300 *translucens* have acted as the founder population in the other countries (Figures S4, S5).
301 Several unique MHs were determined among the Iranian strains, highlighting the allelic
302 richness of the *X. translucens* population in Iran. All these evidences support the hypothesis
303 that a given plant pathogen possess higher genetic diversity in the center of origin of its host
304 plant which is the case in BLS pathogen. Results of neutrality tests (*i.e.* Fu and Li's D^* and Fu
305 and Li's F^*) in the *rpoD* gene sequences of non-Iran *X. translucens* pv. *undulosa* strains also
306 support the hypothesis suggesting a selective sweep and population expansion after a recent
307 bottleneck in the North America, while additional housekeeping genes sequences are needed
308 to be examined to further validate the hypothesis (Table 2). Since the center of origin of
309 cultivated wheat was determined in southeast Turkey (10, 11), we speculate that *X.*
310 *translucens* should be an endemic pathogen in the Middle Eastern countries including Iran.
311 Unlike modern crops such as tomato (*Solanum lycopersicum*) which have only a short history
312 of cultivation and trade in the global-scale backing to the early 16th century (18), the history
313 of worldwide distribution of common wheat by the human kind is not clear, making the
314 phylogeographic studies of cereal pathogens complicated. BLS was reported in the neighbor
315 countries of Iran; hence, a comprehensive epidemiological investigation will decipher the
316 distribution, population structure, and genetic diversity of *X. translucens* in the center of
317 origin of cultivated wheat.

318 According to the eBURST and POPArt analyses, MH2 − including 56 out of 178 strains – was
319 determined as the founder genotype of *X. translucens* pv. *undulosa* (Figure S5). Interestingly,
320 the pathotype strains of *X. translucens* pv. *undulosa* (LMG 892^T) and *X. translucens* pv. *secalis*
321 (LMG 883^T) (both were isolated in southern Canada in 1966) belong to MH2 (Table S1).

322 Pathotype strain of *X. translucens* pv. *undulosa* was isolated from *Triticum turgidum*, while
323 the pathotype strain of *X. translucens* pv. *secalis* was isolated from rye (6). It is more likely
324 that the two pathotypes originated from the same founding genotype in the North America,
325 however, the prevalence of the "one host–one species" concept in the mid-20th century led to
326 the designation of the two strains as two different taxa at that time (4). *Triticum turgidum*,
327 from which the pathotype strain of *X. translucens* pv. *undulosa* was isolated, is also known as
328 "Khorasan wheat" (19). The common name refers to "Khorasan province" in northeastern
329 Iran and indicates the origin of the crop in the Iranian Plateau or Fertile Crescent in the
330 ancient civilizations (19). Khorasan wheat has accidentally been introduced into the USA in
331 1949 and planted for the first time in Fort Benton Montana; while the extended cultivation of
332 the crop has only been started in 1964 in the country (19). Since the founding genotype of the
333 pathogen is still prevalent in the Iranian wheat cultivating areas, one can hypothesize that *X.*
334 *translucens* pv. *undulosa* has been transmitted from the Iranian Plateau into the USA via
335 infected seeds of Khorasan wheat in the mid-20th century. All these evidences bring us to the
336 speculation that the existing population of *X. translucens* pv. *undulosa* in North America was
337 introduced into the "New World" from the Iranian Plateau, and dispersed throughout the
338 continent during the past 50 years (Figure S4; 6). However, further genotypic and
339 phylogeographic examinations using a broader set of strains needed to validate this
340 hypothesis.

341 In conclusion, results of the present study revealed a greater diversity among the worldwide
342 population of *X. translucens* than the one that had been described before. More specifically,
343 based on the phylogeographic analyses we have noted that the strains isolated in the Iranian

344 Plateau – where is considered the center of origin of the cultivated wheat – were genetically
345 more diverse than those of the strains isolated in the New World. These results suggest that
346 the phylogeny of wheat- and barley-pathogenic members of *X. translucens* should be
347 reexamined using a larger collection of strains from all the known hosts of the pathogen,
348 including gramineous weeds, from all over the world. Moreover, despite a few preliminary
349 studies (20–23), these findings raise questions about whether there is any source of
350 resistance among the wild population of wheat species (*Triticum* spp.) in the center of origin
351 of this crop, and emphasize at the same time for a more detailed investigation in this regard.
352 Only future studies — based on population genetics, comparative genomics, and
353 pathogenicity assays of a wide collection of strains isolated from different geographical
354 regions — can shed more light on these questions.

355

356 **Materials and Methods**

357 **Bacterial Strains and Their Identification**

358 Bacterial strains used in this study were obtained during a decade-long (2008-2017) field
359 surveys across cereal-growing areas of northwestern, western, central, and southern Iran.
360 Furthermore, seed lots of small-grain cereals deposited in the seed banks of Seed and Plant
361 Certification and Registration Institute (Karaj, Iran) were investigated for the BLS infestation.
362 Surveying strategy and sampling procedure were the same as described previously (24–25).
363 Bacterial strains were isolated from the plant tissues and seed samples using the methods
364 described previously (26) with minor modifications (27). All the strains (Table 1) were

365 subjected to preliminary phenotypic tests (26). Gram reaction and colony characteristics on
366 yeast extract-dextrose-calcium carbonate (YDC) agar medium were determined.
367 hypersensitivity reaction (HR) was evaluated on tobacco (*Nicotiana tabacum* cv. Turkish)
368 leaves using the bacterial suspension from a 48-hour old culture on NA medium at a
369 concentration of 10^8 CFU/ml. Reference strains of *X. translucens* pv. *undulosa* (ICMP 11055)
370 and *X. translucens* pv. *translucens* (DSM 18974^T) were included as controls in all the tests.
371 Phenotypic tests were repeated twice.

372 **Pathogenicity Tests and Host Range Assays**

373 All the bacterial strains (Table 1) were evaluated for their pathogenicity on six gramineous
374 species including their host of isolation, as well as common gramineous weeds in the region.
375 Barley (cv. Valfajr), Harding's grass, oat, rye, ryegrass, and wheat (cv. Pishtaz) plants were
376 inoculated with the bacterial suspensions using the procedure as described previously (27).
377 The inoculated plants were maintained in greenhouse conditions with 90-95% relative
378 humidity and periodically (daily) monitored for symptom development until 25 days post
379 inoculation (dpi). The same number of plants were treated with the reference strains of *X.*
380 *translucens* pv. *translucens* (DSM 18974^T) and *X. translucens* pv. *undulosa* (ICMP 11055) as
381 positive controls. Negative control plants were treated in the same manner while sterile
382 distilled water was used instead of the bacterial suspensions. Koch's postulates were
383 accomplished by re-isolating the inoculated strains on YDC medium, from plants showing
384 symptoms. Confirmation of the re-isolated bacteria was made using *X. translucens*-specific
385 PCR primers T1/T2 (Table 3; 7). The pathogenicity tests were repeated twice.

386 **Molecular Phylogenetic Analysis**

387 All strains showing xanthomonad-like characteristics (*i.e.* pale-yellow mucoidal colonies on
388 YDC medium and hypersensitivity reaction on tobacco leaves) were subjected to the
389 diagnostic PCR test using the *X. translucens*-specific primer pair T1/T2 (Table 3; 7). Bacterial
390 DNA extraction and PCR procedure were described previously (27), while the primer
391 sequences and annealing temperatures are provided in Table 3. Sixty-five *X. translucens*
392 strains were identified using the species-specific PCR primer and subjected to the MLSA/MLST
393 analyses using the sequences of four housekeeping genes (*i.e. dnaK*, *fyuA*, *gyrB*, and *rpoD*), as
394 recommended previously (6).

395 To determine the phylogenetic position of the strains isolated in Iran, and access a global-
396 scale phylogeography on the BLS pathogen, sequences of the four housekeeping genes in a
397 collection of 112 worldwide *X. translucens* strains were retrieved from the GenBank database
398 and included in the phylogenetic analyses (Table S1; 6, 14). The sequences were concatenated
399 following the alphabetic order of the genes, ending in a sequence of 2464 bp: nucleotides 1-
400 763 for *dnaK*, 764-1285 for *fyuA* (522 bp), 1286-1790 for *gyrB* (505 bp), and 1791-2464 for
401 *rpoD* (674 bp) genes. Phylogenetic trees were constructed using MrBayes 3.1.0 with running
402 the chain for 1×10^6 generations (bootstrapping at 1000 replications) and setting the burnin
403 at 1000 (28). The substitution model for Bayesian analysis was determined using
404 MrModelTest v. 2.3 software as AIC + [GTR+I] (29). The resulted phylogenetic trees were
405 annotated and presented using MEGA 6.06 software (30).

406 **Genetic Diversity and Recombination Analyses**

Nucleotide diversity, number of STs and MHs, haplotype diversity, haplotype frequency, percentage of polymorphic sites, minimum number of recombination events, and the ratio of the rate of non-synonymous to synonymous mutations were estimated using DnaSP 5.10 software (31). The class I neutrality tests (Tajima's *D*, Fu and Li's D^* and Fu and Li's F^* statistics) were also calculated for detecting departure from the mutation/drift equilibrium (31). Detection of potential recombinant sequences and identification of likely parental sequences were carried out using a set of seven non-parametric detection methods (*i.e.* RDP, Geneconv, MaxChi, Chimaera, BootScan, SiScan, and 3Seq) implemented in Recombination Detection Program (RDP) version 4.80 (32). The analysis was performed with default settings for the different detection methods, and the Bonferroni-corrected *P* value cutoff was set at 0.05. Recombination events were accepted when they were identified by at least four out of seven detection methods (32). Splits-decomposition network was constructed and the pairwise homoplasy index (PHI) was calculated using SplitsTree version 4.14.4 (33). These calculations were performed once for 66 Iranian strains, and again for the entire data set (178 strains) for all the individual genes, as well as for the concatenated sequences (33).

Phylogeographic Analyses

To visualize the phylogeographic relationships between the strains, haplotype networks were inferred for individual housekeeping genes using the TCS algorithm (34) implemented in PopART v. 1.7 program (35). Geographic distribution of 178 strains collected in seven countries (*i.e.* Australia, Canada, India, Iran, Paraguay, Mexico and the USA) was employed into the haplotype network of each of the four housekeeping genes, as well as concatenated sequences as described by Leigh and Bryant (35). The country-specific STs, as well as the

shared MHs between the countries were estimated based on the resulting haplotype networks (Table S1). Furthermore, a global-scale geographical structure of *X. translucens* was visualized using the eBURST software (36). The eBURST was used to explore the patterns of evolutionary descent among the strains used in this study as described by Feil et al. (36). The analyses were performed using the stringent (default) group definition, in which sequence types are included within the same clonal complex only if they share identical alleles at three or four out of the four MLST loci with at least one other ST in the population. Furthermore, as a default setting, 1000 re-samplings were performed for bootstrapping the resulted evolutionary network.

GenBank Accession Numbers

The nucleotide sequences were deposited in the GenBank database with the following accession numbers: MK622011−MK622075 for *dnaK*, MK622076−MK622140 for *fyuA*, MK622141−MK622205 for *gyrB* and MK622206−MK622270 for *rpoD* sequences.

Author Contributions

EO and MK conceived and designed the study, with assistance from MT, HH, PK and RK. MK carried out the experiments. EO analyzed and interpreted the data with assistance from MK, HH, RK, and CB. MK and EO prepared the paper, with assistance from CB, RK, and GC. EO revised the final manuscript and acted as the corresponding author.

449 **Conflict of Interest**

450 The authors declare no competing financial interests.

451

452 **Data Accessibility**

453 The data that support the findings of this study are available in the NCBI GenBank database.

454 **Acknowledgments**

455 Financial support for this study was co-provided by Shiraz University and Vali-e-Asr University
456 of Rafsanjan, Iran. CB, RK and EO benefited from interactions promoted by COST Action
457 CA16107 EuroXanth (https://euroxanth.eu/), supported by COST (European Cooperation in
458 Science and Technology).

459

460 **References**

461 1. Jones LR, Johnson AG, Reddy CS. 1917. Bacterial-blight of barley. J Agric Res 11:625-643.
462 2. Duveiller E, Fucikovsky L, Rudolph K. 1997. The Bacterial Diseases of Wheat: Concepts and
463 Methods of Disease Management. CIMMYT. Mexico City.
464 3. Langlois PA, Snelling J, Hamilton JP, Bragard C, Koebnik R, Verdier V, Triplett LR, Blom J,
465 Tisserat NA, Leach JE. 2017. Characterization of the *Xanthomonas translucens* complex
466 using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic
467 LAMP assays. Phytopathology 107:519-527.

4. Vauterin L, Hoste B, Kersters K, Swings J. 1995. Reclassification of *Xanthomonas*. Int J Syst Evol Microbiol 45:472-489.

5. Bragard C, Singer E, Alizadeh A, Vauterin L, Maraite H, Swings J. 1997. *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. Phytopathology 87:1111-1117.

6. Curland RD, Gao L, Bull CT, Vinatzer B, Dill-Macky R, Von Eck L, Ishimaru C. 2018. Genetic diversity and virulence of wheat and barley strains of *Xanthomonas translucens* from the Upper Midwestern United States. Phytopathology 108:443-453.

7. Maes M, Garbeva P, Kamoen O. 1996. Recognition and detection in seed of the *Xanthomonas* pathogens that cause cereal leaf streak using rDNA spacer sequences and polymerase chain reaction. Phytopathology 86:63-69.

8. Schaad NW, Forster RL. 1985. A semiselective agar medium for isolating *Xanthomonas campestris* pv. *translucens* from wheat seeds. Phytopathology 75:260-263.

9. Rademaker JLW, Norman DJ, Forster RL, Louws FJ, Schultz MH, De Bruijn FJ. 2006. Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. Phytopathology 96:876-884.

10. Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278: 1312-1314.

11. Brandolini A, Volante A, Heun M. 2016. Geographic differentiation of domesticated einkorn wheat and possible Neolithic migration routes. Heredity 117: 135-141.

12. Alizadeh A, Rahimian H. 1989. Bacterial leaf streak of *Gramineae* in Iran. Bull OEPP *19*:113-117.

13. FAOSTAT (2018). The agriculture production domain. Available at http://www.fao.org/faostat/en/?#compare

14. Peng Z, Hu Y, Xie J, Potnis N, Akhunova A, Jones J, Liu Z, White FF, Liu S. 2016. Long read and single molecule DNA sequencing simplifies genome assembly and TAL effector gene analysis of *Xanthomonas translucens*. BMC Genomics 17:21-39.

15. Osdaghi E, Taghavi SM, Hamzehzarghani H, Fazliarab A, Lamichhane JR. 2017. Monitoring the occurrence of tomato bacterial spot and range of the causal agent *Xanthomonas perforans* in Iran. Plant Pathol 66:990-1002.

16. Osdaghi E 2014. Occurrence of common bacterial blight on mungbean (*Vigna radiata*) in Iran caused by *Xanthomonas axonopodis* pv. *phaseoli*. New Disease Reports, 30: 9.

17. Mafakheri H, Taghavi SM, Banihashemi Z, Osdaghi E, Lamichhane JR 2017. Pathogenicity, host range and phylogenetic position of *Agrobacterium* species associated with sugar beet crown gall outbreaks in Southern Iran. Eur J Plant Pathol, 147: 721-730.

18. Peralta IE, Spooner DM, Razdan MK, Mattoo AK. 2006. History, origin and early cultivation of tomato (Solanaceae), p 1 -27. In Razdan MK, Mattoo AK (ed), Genetic Improvement of Solanaceous Crops, 1st ed, vol 2. CRC Press, Boca Raton, FL.

19. Sacks G. 2005. Kamut: A New Old Grain. Gastronomica 5:95-98.

20. Duveiller E, van Ginkel M, Thijssen M. 1992. Genetic analysis of resistance to bacterial leaf streak caused by *Xanthomonas campestris* pv. *undulosa* in bread wheat. Euphytica 66:35-43.

511 21. Adhikari TB, Hansen JM, Gurung S, Bonman JM. 2011. Identification of new sources of
512 resistance in winter wheat to multiple strains of *Xanthomonas translucens* pv.
513 *undulosa*. Plant Dis 95:582-588.
514 22. Kandel YR, Glover KD, Tande CA, Osborne LE. 2012. Evaluation of spring wheat germplasm
515 for resistance to bacterial leaf streak caused by *Xanthomonas campestris* pv.
516 *translucens*. Plant Dis 96:1743-1748.
517 23. Wen A, Jayawardana M, Fiedler J, Sapkota S, Shi G, Peng Z, Liu S, White FF, Bogdanove AJ,
518 Li X, Liu Z. 2018. Genetic mapping of a major gene in triticale conferring resistance to
519 bacterial leaf streak. Theor Appl Genet 131:649-658.
520 24. Osdaghi E, Taghavi SM, Fazliarab A, Elahifard E, Lamichhane JR (2015). Characterization,
521 geographic distribution and host range of *Curtobacterium flaccumfaciens*: An
522 emerging bacterial pathogen in Iran. Crop Prot, 78: 185-192.
523 25. Osdaghi E, Taghavi SM, Hamzehzarghani H, Fazliarab A, Harveson RM, Lamichhane JR
524 2016. Occurrence and characterization of a new red-pigmented variant of
525 *Curtobacterium flaccumfaciens*, the causal agent of bacterial wilt of edible dry beans
526 in Iran. European Eur J Plant Pathol, 146: 129-145.
527 26. Schaad NW, Jones JB, Chun W. 2001. Laboratory guide for the identification of plant
528 pathogenic bacteria, 3th ed. APS Press.
529 27. Osdaghi E, Ansari M, Taghavi SM, Zarei S, Koebnik R, Lamichhane JR (2018. Pathogenicity
530 and phylogenetic analysis of *Clavibacter michiganensis* strains associated with tomato
531 plants in Iran. Plant Pathol, 67:957-970.

532 28. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,
533 Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic
534 inference and model choice across a large model space. Syst Biol 61:539-542.
535 29. Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey J. 2004. Bayesian phylogenetic
536 analysis of combined data. Syst Biol 53:47-67.
537 30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular
538 evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725-2729.
539 31. Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA
540 polymorphism data. Bioinformatics 25:1451-1452.
541 32. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis
542 of recombination patterns in virus genomes. Virus Evol 1:vev003.
543 33. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies.
544 Mol Biol Evol 23:254-267.
545 34. Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene
546 genealogies. Mol Ecol 9:1657-1659.
547 35. Leigh JW, Bryant D. 2015. Popart: full-feature software for haplotype network
548 construction. Methods Ecol Evol 6:1110-1116.
549 36. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of
550 evolutionary descent among clusters of related bacterial genotypes from multilocus
551 sequence typing data. J Bacteriol 186:1518-1530.

552 37. Young, J. M., Park, D. C., Shearman, H.M., & Fargier, E. (2008). A multilocus sequence
553 analysis of the genus *Xanthomonas*. Systematic and Applied Microbiology, 31(5), 366-
554 377.
555 38. Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., & Stead, D. (2007). Phylogenetic
556 analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences.
557 International Journal of Systematic and Evolutionary Microbiology, 57(12), 2881-2887.

558

559

560 **Legends for Figures and Tables**

561 **Figure 1**: Phylogeny of *Xanthomonas translucens* strains isolated in Iran based on the
562 concatenated sequences of four housekeeping genes (*i.e. dnaK, fyuA, gyrB,* and *rpoD*).
563 *Xanthomonas translucens* pv. *poae* (LMG 728^{T}) was used as an out-group to root the
564 tree. The bar presents numbers of substitutions per site. Among the 65 Iranian strains,
565 57 strains were identified as *X. translucens* pv. *undulosa* while eight strains were
566 identified as *X. translucens* pv. *translucens*. Black and white table in the rightmost side
567 of the figure shows host range of the strains on six gramineous species including
568 wheat (W), barley (B), oat (O), rye (R), ryegrass (Rg), and Harding's grass (P: *Phalaris*)
569 where black squares indicate pathogenicity and white squares indicate non-
570 pathogenicity. The atypical strain XtKm18 (indicated with asterisk inside the black
571 square), which was isolated from barley and identified as *X. translucens* pv.
572 *translucens*, induced water-soaking symptoms on wheat plants. Blue background

573 indicates *X. translucens* pv. *undulosa* strains, while red, yellow, and green backgrounds
574 indicate different sub-clusters of *X. translucens* pv. *translucens*.

575 **Figure 2**: TCS multilocus haplotype (MH) network generated using the POPArt program from
576 the concatenated partial sequences of four housekeeping genes (*i.e. dnaK*, *fyuA*, *gyrB*,
577 and *rpoD*) in 178 *Xanthomonas translucens* strains representing strains from seven
578 countries around the globe. The size of the circles indicates the relative frequency of
579 sequences belonging to a particular MH. Hatch marks along the branches indicate the
580 number of mutations. Each color indicates a different geographic area. To provide a
581 precise insight into the diversity of the pathogen, strains isolated in southern Iran and
582 western Iran were considered distinct groups. The numbers in the left side of the radix
583 character represent the MH number according to Table S1, while the numbers in the
584 right side of the radix character indicate the number of strains in a given MH.
585 *Xanthomonas translucens* pv. *undulosa* strains are surrounded by blue lines while *X.*
586 *translucens* pv. *translucens* strains are surrounded by red lines. MH2 consisting of 56
587 *X. translucens* pv. *undulosa* strains was considered the founder genotype of the
588 pathovar as confirmed by the phylogeographic analyses using the eBURST program
589 (Figure S5).

590 **Figure 3**: TCS haplotype network generated using the POPArt program from the partial
591 sequences of four housekeeping genes (*i.e. dnaK*, *fyuA*, *gyrB*, and *rpoD*) in 178
592 *Xanthomonas translucens* strains. The size of the circles indicates the relative
593 frequency of sequences belonging to a particular sequence type. Hatch marks along
594 the branches indicate the number of mutations. Each color indicates a different

595 geographic area. To provide a precise insight into the diversity of the pathogen, strains
596 isolated in southern Iran and western Iran were considered distinct groups. The
597 numbers in the left side of the radix character represent the ST number according to
598 Table S1, while the numbers in the right side of the radix character indicate the
599 number of strains in a given ST. Strains isolated in Iran generated several unique
600 haplotypes, indicating a higher genetic diversity of the pathogen in Iran.

601 **Table 1:** *Xanthomonas translucens* strains isolated in Iran, their host plant, geographic area
602 and year of isolation, as well as their sequence types based on the partial sequences of
603 four housekeeping genes (*i.e. dnaK, fyuA, gyrB*, and *rpoD*).

604 **Table 2:** Sequence variation statistics of the partial sequences of four housekeeping genes
605 (*i.e. dnaK, fyuA, gyrB*, and *rpoD*) from *Xanthomonas translucens* strains isolated in Iran
606 in comparison to the strains isolated in six other countries.

607 **Table 3:** Primer pairs used in this study.

608 **Figure S1:** Distribution if *Xanthomonas translucens* strains isolated in this study across the
609 wheat and barley growing areas in Iran. Each bacterial strain in Table 1 is represented
610 by one circle on the map. Yellow circles indicate *X. translucens* pv. *undulosa* strains
611 while red circles indicate *X. translucens* pv. *translucens* strains.

612 **Figure S2:** Phylogenetic tree generated from the concatenated partial sequences of four
613 housekeeping genes (*i.e. dnaK, fyuA, gyrB*, and *rpoD*) in 178 *Xanthomonas translucens*
614 strains isolated from different countries, as well as the type strains of the
615 phylogenetically closely relates species/pathovars. Black circles indicate the strains

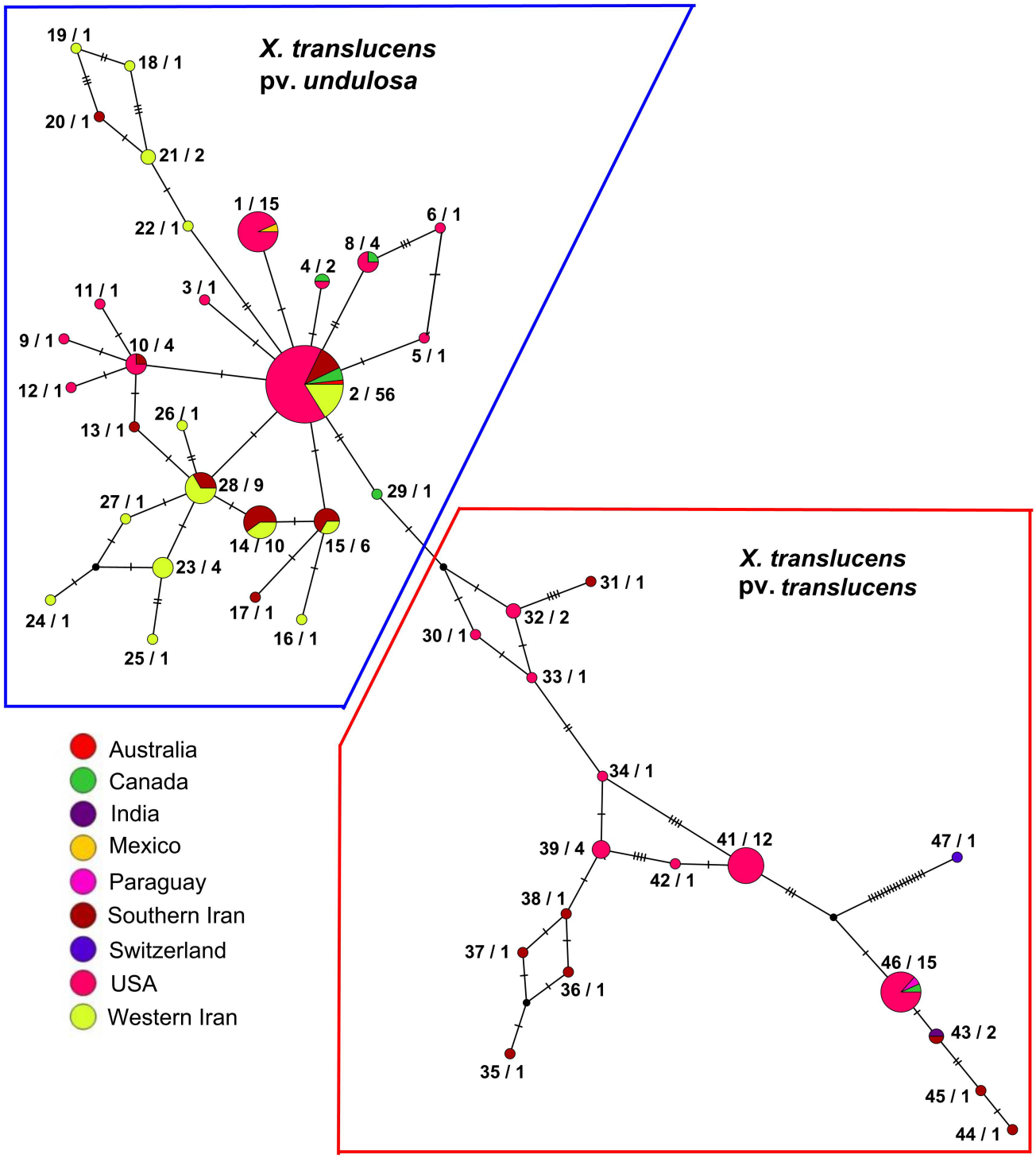
isolated in Iran. Blue background indicates *X. translucens* pv. *undulosa* strains while red background indicates *X. translucens* pv. *translucens* strains. All the *X. translucens* pv. *undulosa* and *X. translucens* pv. *translucens* strains were clustered in a monophyletic clade.

Figure S3: TCS haplotype network generated using POPArt program from the partial sequences of four housekeeping genes *dnaK, fyuA, gyrB*, and *rpoD* in 66 *Xanthomonas translucens* strains isolated in Iran. The size of the circles indicates the relative frequency of sequences belonging to a particular haplotype. Hatch marks along the branches indicate the number of mutations. Each color indicates a different geographic location (province) in Iran. *X. translucens* pv. *translucens* strains possessing unique ST and differing from the *X. translucens* pv. *undulosa* strains are framed by a grey box.

Figure S4: Worldwide distribution of *Xanthomonas translucens*, the causal agent of bacterial leaf streak on small-grain cereals. The center of origin of cultivated wheat in southern Turkey is marked by yellow concentric circles. Dashed arrows show the hypothetical transmission routes of *X. translucens* across the nations. TCS networks inside the zoomed-in map of Iran and the USA show MHs of the bacterial strains isolated in each country. Diversity parameters (*i.e.* higher haplotype frequency and richness) indicate the higher genetic diversity of X. *translucens* in Iran. N: number of strains; HF: haplotype frequency.

636 **Figure S5:** Analysis of worldwide collection of 178 *Xanthomonas translucens* strains using
637 eBURST with the stringent (default) group definition in which, MHs are included within
638 the same clonal complex only if they share identical alleles at three or four of the four
639 MLST loci with at least one other MH in the population. The predicted primary
640 founder of *X. translucens* pv. *undulosa*, MH2 (bootstrap confidence value of 99%) is
641 shown as a blue circle while the descendent subgroups are shown as yellow circles.
642 Numbers in the eBURST diagram correspond to MHs as shown in Table S1. The
643 diameter of each circle corresponds to the abundance of the strains of the
644 corresponding MH in the input data. *Xanthomonas translucens* pv. *translucens* (CIX34)
645 and *X. translucens* pv. *undulosa* (CIX103) strains, which are single-locus variants, are
646 specified in the diagram.

647 **Table S1:** *Xanthomonas translucens* strains of worldwide origin that were used in this study,
648 their host plant, geographic area and date of isolation, as well as their Multilocus
649 Haplotypes (MHs) and Sequence Types (STs) based on the partial sequences of four
650 housekeeping genes *(i.e. dnaK, fyuA, gyrB,* and *rpoD*).



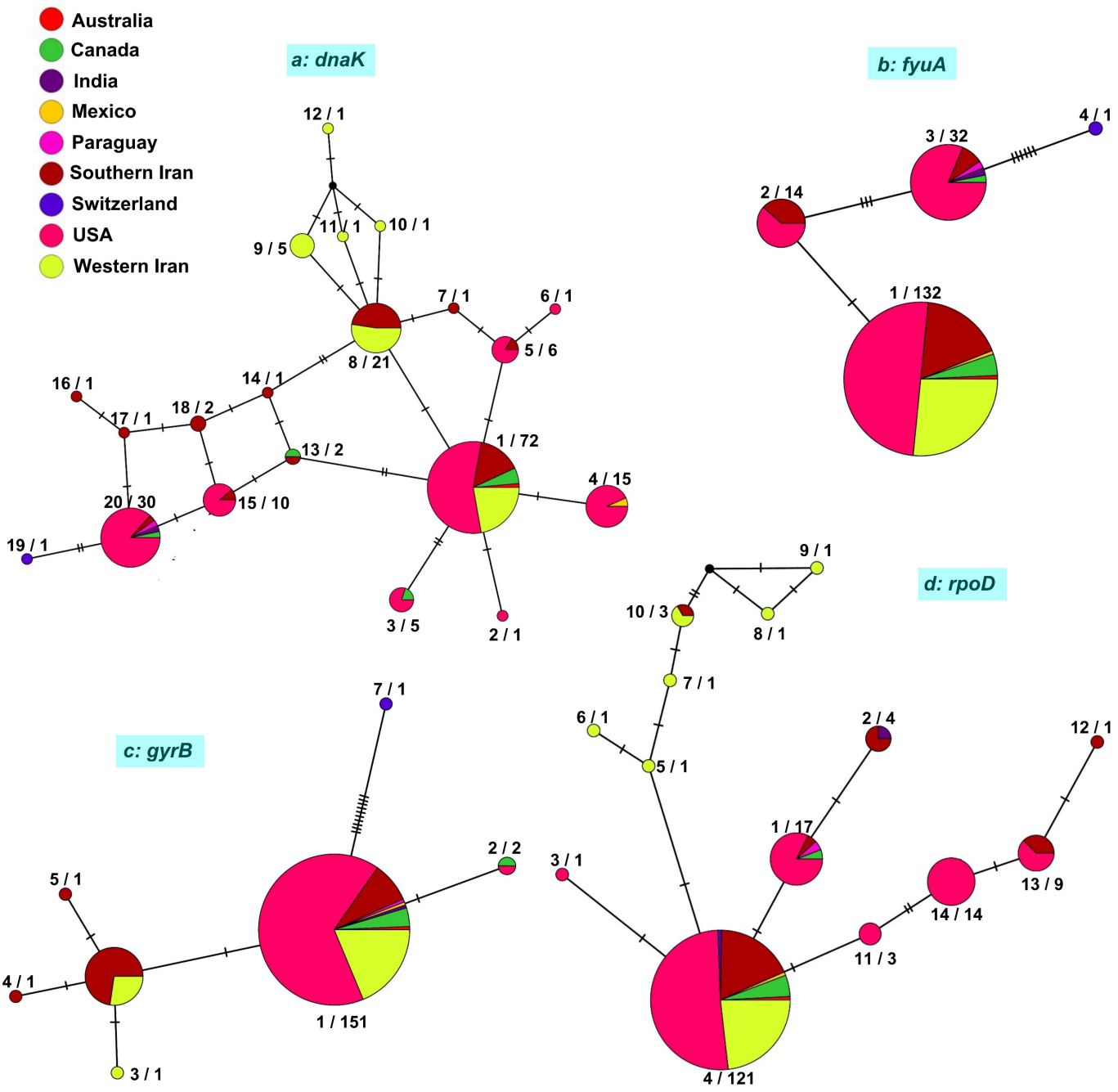


Table 1: *Xanthomonas translucens* strains isolated in Iran, their host plant, geographic area and year of isolation, as well as their sequence types based on the partial sequences of four housekeeping genes (*i.e. dnaK, fyuA, gyrB,* and *rpoD*).

Strain	Host	Region		Year	Sequence types			
		Province	County		*dnaK*	*fyuA*	*gyrB*	*rpoD*
XtLr1	Wheat	Lorestan	Borujerd	2017	8	1	6	4
XtLr2	Wheat	Lorestan	Borujerd	2017	9	1	1	4
XtLr3	Wheat	Lorestan	Borujerd	2017	1	1	6	4
XtLr4	Wheat	Lorestan	Khorramabad	2016	1	1	1	4
XtLr5	Wheat	Lorestan	Khorramabad	2016	10	1	1	5
XtLr6	Wheat	Lorestan	Khorramabad	2016	1	1	1	8
XtLr7	Barley	Lorestan	Khorramabad	2017	1	1	1	10
XtLr8	Wheat	Lorestan	Khorramabad	2016	9	1	1	6
XtLr9	Barley	Lorestan	Khorramabad	2017	1	1	1	10
XtKr1	Wheat	Kermanshah	Kermanshah	2016	8	1	6	4
XtKr2	Wheat	Kermanshah	Kermanshah	2016	9	1	1	4
XtKr3	Wheat	Kermanshah	Kermanshah	2016	1	1	3	4
XtKr4	Wheat	Kermanshah	Kermanshah	2016	11	1	1	4
XtKr5	Wheat	Kermanshah	Kermanshah	2016	1	1	1	4
XtZa1	Wheat	Zanjan	Khodabandeh	2016	8	1	6	4
XtZa2	Wheat	Zanjan	Khodabandeh	2016	9	1	1	4
XtZa3	Wheat	Zanjan	Khodabandeh	2016	8	1	1	4
XtZa4	Wheat	Zanjan	Khodabandeh	2016	8	1	1	9
XtZa5	Wheat	Zanjan	Khodabandeh	2016	1	1	1	7
XtFa1	Wheat	Fars	Firuzabad	2016	8	1	1	4
XtFa2	Wheat	Fars	Eqlid	2016	8	1	6	4
XtFa3	Wheat	Fars	Eqlid	2016	8	1	6	4
XtFa4	Wheat	Fars	Firuzabad	2016	8	1	1	10
XtFa5	Wheat	Fars	Eqlid	2016	8	1	1	4
XtFa6	Wheat	Fars	Eqlid	2016	8	1	1	4
XtFa7	Wheat	Fars	Eqlid	2016	1	1	1	4
XtKn1	Wheat	Kurdistan	Dehgolan	2016	1	1	6	4

XtKn2	Wheat	Kurdistan	Dehgolan	2016	8	1	1	4
XtKn3	Wheat	Kurdistan	Dehgolan	2016	1	1	1	4
XtKn4	Wheat	Kurdistan	Marivan	2016	1	1	1	4
XtKn5	Wheat	Kurdistan	Marivan	2016	1	1	1	4
XtHn1	Wheat	Hamadan	Nahavand	2017	8	1	1	4
XtHn2	Wheat	Hamadan	Nahavand	2017	8	1	6	4
XtHn3	Wheat	Hamadan	Nahavand	2017	1	1	1	4
XtHn4	Wheat	Hamadan	Asadabad	2016	1	1	1	4
XtHn5	Wheat	Hamadan	Asadabad	2016	1	1	1	4
XtHn6	Wheat	Hamadan	Asadabad	2016	8	1	1	4
XtHn7	Wheat	Hamadan	Asadabad	2016	12	1	1	4
XtHn8	Wheat	Hamadan	Asadabad	2016	9	1	1	4
XtHn9	Barley	Hamadan	Asadabad	2017	8	1	1	4
XtHn10	Barley	Hamadan	Asadabad	2017	1	1	1	4
XtHn11	Barley	Hamadan	Asadabad	2017	8	1	1	4
XtKm3	Rye	Kerman	Bardsir	2015	5	1	1	4
XtKm4	Barley	Kerman	Bardsir	2015	16	3	6	2
XtKm7	Barley	Kerman	Shahr-e Babak	2014	18	2	4	1
XtKm8	Barley	Kerman	Shahr-e Babak	2014	15	2	6	13
XtKm9	Barley	Kerman	Bardsir	2015	20	3	1	2
XtKm10	Barley	Kerman	Bardsir	2015	8	1	6	4
XtKm11	Barley	Kerman	Bardsir	2015	17	3	6	2
XtKm12	Wheat	Kerman	Bardsir	2015	7	1	1	4
XtKm15	Ryegrass	Kerman	Orzuiyeh	2015	1	1	6	4
XtKm18	Barley	Kerman	Baft	2015	18	2	6	13
XtKm22	Wheat	Kerman	Kerman	2008	8	1	6	4
XtKm24	Wheat	Kerman	Kerman	2008	8	1	6	4
XtKm25	Wheat	Kerman	Kerman	2015	1	1	6	4
XtKm27	Wheat	Kerman	Kerman	2015	1	1	6	4
XtKm28	Wheat	Kerman	Kerman	2009	1	1	1	4
XtKm29	Wheat	Kerman	Kerman	2015	1	1	1	4
XtKm30	Wheat	Kerman	Kerman	2015	8	1	6	4
XtKm31	Wheat	Kerman	Kerman	2015	1	1	5	4
XtKm33	Barley	Kerman	Orzuiyeh	2009	14	2	6	12
XtKm34	Barley	Kerman	Kerman	2015	13	2	6	13
XtKm35	Wheat	Kerman	Kerman	2015	1	1	6	4
XtBu36	Wheat	Bushehr	Borazjan	2017	1	1	1	4

XtBu38	Wheat	Bushehr	Borazjan	2017	1	1	1	4
ICMP 11055	Wheat	Kerman	Kerman	1983	1	1	1	4
ICMP 5752	Barley	USA		ND	21	2	1	13

Table 2: Sequence variation statistics of the partial sequences of four housekeeping genes (*i.e. dnaK, fyuA, gyrB,* and *rpoD*) from *Xanthomonas translucens* strains isolated in Iran in comparison to the strains isolated in six other countries.

Region		Gene	NS	N	H	HF	S	Pol (%)	P_i (π)	NM	HD	Neutrality tests			R
												Tajima's *D*	Fu and Li's *D**	Fu and Li's *F**	
Iran	*Xtu*	***dnaK***	58[1]	762	8	0.137	5	0.656	0.118×10^{-2}	5	0.655	-0.391 NS	1.083 NS	0.724 NS	2
		fyuA	58	522	1	0.017	0	0.000	0.000	0	0.000	ND	ND	ND	0
		gyrB	58	505	4	0.068	3	0.594	0.100×10^{-2}	3	0.456	-0.442 NS	-1.782 NS	-1.602 NS	0
		rpoD	58	674	7	0.120	6	0.890	0.123×10^{-2}	7	0.257	-1.165 NS	-0.369 NS	-0.737 NS	2
	Xtt	***dnaK***	8	762	7	0.875	4	0.524	0.230×10^{-2}	4	0.964	0.586 NS	0.568 NS	0.629 NS	1
		fyuA	8	522	2	0.250	3	0.574	0.308×10^{-2}	3	0.536	1.600 NS	1.233 NS	1.445 NS	0
		gyrB	8	505	3	0.375	2	0.396	0.099×10^{-2}	2	0.464	-1.310 NS	-1.409 NS	-1.513 NS	0
		rpoD	8	674	4	0.500	7	1.038	0.540×10^{-2}	7	0.786	1.662 NS	0.973 NS	1.252 NS	0
Non-Iran	*Xtu*	***dnaK***	75	762	8	0.106	8	1.049	0.129×10^{-2}	8	0.598	-1.011 NS	0.512 NS	0.004 NS	1
		fyuA	75	522	2	0.026	1	0.191	0.005×10^{-2}	1	0.027	-1.062 NS	-1.947 NS	-1.958 NS	0
		gyrB	75	505	2	0.026	1	0.198	0.010×10^{-2}	1	0.053	-0.904 NS	0.513 NS	0.110 NS	0
		rpoD	74*	674	5	0.067	5	0.741	0.031×10^{-2}	5	0.155	-1.780 NS	-3.024*, P<0.05	-3.084*, P<0.05	0
	Xtt	***dnaK***	36*	762	2	0.055	1	0.131	0.042×10^{-2}	1	0.322	0.495 NS	0.574 NS	0.636 NS	0
		fyuA	37	522	2	0.054	3	0.574	0.200×10^{-2}	3	0.348	1.009 NS	0.924 NS	1.102 NS	0
		gyrB	37	505	1	0.027	0	0.000	0.000	0	0.000	ND	ND	ND	ND
		rpoD	37	674	5	0.135	6	0.890	0.353×10^{-2}	6	0.701	1.777 NS	0.386 NS	0.949 NS	0

1: 57 strains isolated in this study and the reference strain ICMP 11055 isolated in southern Iran in 1983.

NS: Number of strains
N: Number of nucleotides
H: Number of haplotypes
HF: Haplotype frequency: ratio of H per NS
S: Total number of segregating sites
Pol: Percentage of polymorphic sites
P_i: Nucleotide diversity (π)
NM: Number of mutations (η)

HD: Haplotype (gene) diversity
R: Minimum number of recombination events
n.p.: no polymorphism
NS: not significant
*: The *dnaK* and *rpoD* genes sequences in ICMP 5752 and CIX127 strains, respectively, were excluded from the analyses due to the unsatisfactory sequence quality.

Table 3: Primer pairs used in this study.

Primer name	Nucleotide sequence (5´- 3´)	Size of amplicon (bp)	Annealing temperature (°C)	Target region	Reference
T1	CCGCCATAGGGCGGAGCACCCCGAT	139	53	16S-23S rRNA	7
T2	GCAGGTGCGACGTTTGCAGAGGGATCTTCTGCAAA				
XrpoD1F	TGGAACAGGGCTATCTGACC	674	54	*rpoD*	37
XrpoD1R	CATTCYAGGTTGGTCTGRTT				
dnaK-F	TCCTAAGCACCTCAACATCAAG	762	59	*dnaK*	6
dnaK-R	CTTCTTGTCGTCCTTGACCTC				
fyuA-F	CTCGCAGAACGGCCTGTA	522	61	*fyuA*	6
fyuA-R	GTAGCCGGGCATCTTCAACT				
XgyrPCR2F	AAGCAGGGCAAGAGCGAGCTGTA	700	54	*gyrB*	38
X.gyrrsp1	CAAGGTGCTGAAGATCTGGTC				