Loss of altruism in the social amoeba *Dictyostelium discoideum* is associated with the G protein-coupled receptor *grlG*

Laura M. Walker[1*], Rintsen N. Sherpa[2], Sindhuri Ivaturi[1], Debra A. Brock[1], Jason R. Walker[3], Joan E. Strassmann[1] and David C. Queller[1]

[1]Department of Biology, Washington University, St. Louis, Missouri 6313

[2]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

[3]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63130

* Corresponding author: Department of Biology, Washington University, 1 Brookings Drive, St. Louis, Missouri 63130; Email: lwalker@wustl.edu

ORCID IDs: 0000-0003-0837-7227 (L.M.W.); 0000-0002-4349-5854 (D.A.B.); 0000-0001-7547-5789 (J.R.W.); 0000-0003-0638-8440 (J.E.S.); 0000-0002-5464-1984 (D.C.Q)

Short title: Loss of altruism in *D. discoideum*

Abstract

Aggregative multicellularity relies on cooperation among individual cells to form a multicellular body. In *Dictyostelium discoideum* this cooperation is maintained by high relatedness. Previous work showed that experimental evolution under low-relatedness resulted in an increase of cheaters (cells that contribute proportionally more to spores than to the sterile stalk) and that many clones completely lost cooperation and the ability to form fruiting bodies. Here, we investigate the genomic changes underlying the evolution of the cheating phenotype using whole-genome sequencing and variant analysis of these previously evolved *D. discoideum* lines. We identified 38 single nucleotide polymorphisms in 29 genes, none of which have been previously implicated in cheating. Each gene has one variant except for the G protein-coupled receptor *grlG*, which has at least one variant in over half of the lines. Upon identifying the parallel evolution of *grlG*, we screened additional clones to investigate the correlation between variants in the gene and the loss of cooperation (identified by the inability to form a fruiting body). We found that variants in the 5' half of *grlG* that impact the signal peptide or extracellular binding domain are significantly associated with the loss of cooperation (non-fruiting); the association was not significant in the 3' half of the gene. This suggests that the loss of *grlG* was adaptive under low-relatedness and that the 5' half of the gene in particular is important for cooperation and multicellular development. This confirms the importance of high relatedness in the evolution of altruism in the social amoeba *D. discoideum.*

Key words: Aggregative multicellularity, *far2*, Glutamate receptor-like protein, Orphan GPCR

Introduction

High relatedness is crucial for the maintenance of cooperation in multicellularity. Without high relatedness to reduce conflict among individuals in a cooperative group, cheaters can evolve that exploit the individual and destabilize altruistic traits (Hardin 1968; Gilbert *et al.* 2007). High relatedness is easily achieved within organisms that go through a single-cell bottleneck or are otherwise clonal, as is true for most of the more than 20 transitions from unicellularity to multicellularity (Grosberg and Strathmann 2007; Knoll 2011). A single-cell origin ensures that all cells are highly related (genetically identical) which reduces conflict, protects against cheaters and promotes the division of labor that is so crucial for complex multicellularity such as in plants and animals (Grosberg and Strathmann 2007; Cooper and West 2018). However, a single-cell origin is not the only way to achieve multicellularity and its many benefits. Many other organisms from across the tree of life achieve multicellularity via aggregation or fusion of individual cells (Bonner 1998; Sebé-Pedrós *et al.* 2017). In this form of multicellularity, called aggregative multicellularity, the cells that aggregate need not be related, so other mechanisms are required to reduce conflict and maintain cooperation (Queller 2000).

Organisms that arise by aggregation offer important opportunities for evolutionary and mechanistic studies of conflict and cooperation. Genetic indications of conflict and its control are likely to be evident in multicellular organisms that arise by aggregation. Unlike multicellular

67 organisms that require a single-cell bottleneck, in organisms with aggregative multicellularity it
68 is possible to experimentally manipulate intra-organismal relatedness and study the outcome.
69 Among the best studied organisms with aggregative multicellularity is the social amoeba
70 *Dictyostelium discoideum.* Usually free-living in the soil, when faced with starvation *D.*
71 *discoideum* amoebae aggregate to cooperatively form a multicellular fruiting body in which
72 about 80% develop into reproductive spores and the remaining 20% die forming a stalk (Kessin
73 2001). The stalk lifts the spores, facilitating dispersal (smith *et al.* 2014). Stalk formation is
74 altruistic because cells that contribute to the stalk die, sacrificing themselves to aid in spore
75 dispersal. This partitioning of reproductive spores and sterile stalk cells provides an opportunity
76 for clones to cheat in mixtures of genotypes by contributing more than their fair share of cells
77 to spores versus the stalk (skewing the typical distribution) (Strassmann *et al.* 2000).
78
79 *Dictyostelium discoideum* is a particularly valuable and tractable model system for studying
80 aggregative multicellularity (Strassmann and Queller 2011; Ostrowski 2019). In the laboratory
81 we can experimentally manipulate populations of genetically distinct amoebae and they will
82 aggregate together and develop into a fruiting body. An extensive set of experimental tools and
83 resources (fluorescent labels, reference genome, etc.) are available that allow us to track the
84 outcome of the distinct genotypes (Buttery *et al.* 2013). In the multicellular fruiting bodies that
85 result from mixing experiments, it is usually possible to identify a winner (the clone contributing
86 more to spores than stalk) and a loser (the clone contributing more to the stalk than spores)
87 (Strassmann *et al.* 2000; Fortunato *et al.* 2003a; Buttery *et al.* 2009; Wolf *et al.* 2015; Madgwick
88 *et al.* 2018). Evidence suggests that cheating occurs in natural populations too. Despite the high

89 relatedness in natural populations (Gilbert *et al.* 2007), distinct clones are often found in close
90 proximity and they too will aggregate and form chimeric fruiting bodies (Strassmann *et al.*
91 2000; Fortunato *et al.* 2003b; smith *et al.* 2016).
92
93 Genetic mutations that result in a selfish trait such as the preferential formation of spores
94 versus stalk (cheater mutations), may pose a threat to cooperation depending on the context in
95 which they arise. Within an individual fruiting body a cheater mutation can be favored by
96 selection and contribute disproportionately more to spores than to stalk (Strassmann *et al.*
97 2000). But the outcome more broadly will depend on the relatedness of aggregating cells and
98 the extent to which those carrying the mutation can still cooperate on their own. If a clone
99 carrying the cheater mutation has completely lost cooperation such that it can only form spores
100 (an obligate cheater), then it can only persist when an exploitable clone is available. When an
101 obligate cheater clone finds itself alone without others to form a stalk, multicellular fruiting
102 body formation (development) comes to a halt resulting in a mound of cells that cannot
103 disperse, and thus has a fitness of zero. We call these obligate cheater clones "non-fruiters." If
104 instead, a clone carrying the cheater mutation can still cooperate and can modify its strategy
105 depending on the availability of exploitable clones then the mutation may persist without
106 threatening the collapse of multicellularity. We call these clones “facultative cheaters.”
107
108 It has been experimentally shown that selection under low relatedness increases the
109 occurrence of both facultative (Santorelli *et al.* 2008) and non-fruiting obligate cheaters (Ennis
110 *et al.* 2000; Kuzdzal-Fick *et al.* 2011). When obligate cheaters arise in the wild where

111 relatedness is high, they should not persist because they will usually find themselves in clonal
112 or nearly clonal aggregations where their cheating advantage will not trump the disadvantage
113 of their inability to form fruiting bodies on their own (Fortunato *et al.* 2003b; Gilbert *et al.*
114 2007). In line with that expectation, no obligate cheaters have been identified from natural
115 populations of *D. discoideum*. While it is true that non-fruiting clones would be missed with
116 traditional isolation techniques (Cavender and Raper 1965), Gilbert *et al.* (2007) screened
117 specifically for the presence of non-fruiting obligate cheaters by clonally plating over 3,000
118 spores from wild-collected fruiting bodies to observe the phenotype. No clones were found that
119 were incapable of forming a fruiting body. These findings are in line with predictions of kin
120 selection theory (Hamilton 1964) and show that high relatedness helps stabilize altruism in *D.*
121 *discoideum.* The altruistic trait of stalk formation can be maintained only if the benefits are
122 shared with related individuals who likely share the altruism allele.
123
124 One way in which organisms with aggregative origins of multicellularity maintain high
125 relatedness and the resulting cooperation is through kin discrimination (West *et al.* 2006). Kin
126 discrimination describes mechanisms that allow individuals to identify and preferentially direct
127 benefits to relatives and either harm or ignore non-relatives (Tsutsui 2004; Strassmann *et al.*
128 2011). Some species of Dictyostelids have robust kin discrimination systems that aid in the
129 sorting of different genotypes into mostly clonal fruiting bodies (Mehdiabadi *et al.* 2006, 2009;
130 Sathe *et al.* 2014). *Dictyostelium discoideum* also has kin discrimination with sorting regulated
131 by two highly variable cell adhesion genes, *tgrB1* and *tgrC1* (Benabentos *et al.* 2009; Hirose *et*
132 *al.* 2011; Gruenheit *et al.* 2017). Segregation by kin discrimination in *D. discoideum* takes place

133 during early aggregation but is not maintained and does not result in fruiting bodies that are
134 highly sorted by relatedness (Flowers *et al.* 2010; Gilbert *et al.* 2012). So although it may limit
135 some cheaters from invading (Ho *et al.* 2013), kin discrimination alone is not strong enough to
136 provide complete protection from selfish cheaters in *D. discoideum*. Instead, high relatedness in
137 natural populations must be maintained in additional ways such as through structured
138 population growth and spatial distribution patterns (Buttery *et al.* 2012; smith *et al.* 2016).
139
140 Despite the great deal has been learned about cooperation and conflict in *D. discoideum,* much
141 remains to be learned with respect to the underlying genes and pathways. Several efforts to
142 identify genes underlying cheating phenotypes have been made, often by employing restriction
143 enzyme-mediated integration (REMI) to generate gene knock-out mutant libraries (Kuspa and
144 Loomis 1992). Screening of such REMI mutant libraries has successfully revealed a number of
145 genes involved in cheating phenotypes in *D. discoideum.* Santorelli *et al.* (2008) screened a
146 large REMI mutant library, focusing on facultative cheater genes (by requiring one round of
147 clonal fruiting body development) and identified over 100 genes predicted to cause cheating
148 when lost to mutation. As might be expected for such a large number of loci, the gene ontology
149 terms span a wide diversity of biological processes. Thus far only two of the genes from
150 Santorelli *et al.* (2008) have been characterized including *chtB* (Santorelli *et al.* 2013) and *chtC*
151 (Khare and Shaulsky 2010). Despite the expectation that mutations in genes that cause cheating
152 should have an associated fitness cost to control their spread, neither the *chtB* nor *chtC*
153 mutants have any obvious fitness costs. Mutant clones for both genes are able to cheat
154 facultatively, increasing in numbers in chimera but when alone, they contribute cells to both

155 spores and stalk and form fruiting bodies normally (Khare and Shaulsky 2010; Santorelli *et al.*
156 2013). It may be that there are disadvantageous pleiotropic effects that keep these mutations
157 from spreading, but they were not detected in the laboratory. Pleiotropy is, after all, another
158 mechanism proposed to hinder the spread of cheaters such as described for two separately
159 identified facultative cheater genes in *D. discoideum*, *dimA* and *csaA* (Queller *et al.* 2003; Foster
160 *et al.* 2004; Strassmann and Queller 2011).
161
162 In addition to the aforementioned facultative cheater genes, only one obligate cheater gene
163 has been described, *fbxA* (Ennis *et al.* 2000). The *fbxA* mutant, also resulting from a REMI
164 mutant library screen, was the first cheater gene to be described and has therefore also been
165 referred to as "Cheater A" or "chtA" for short. When cells from a clone carrying a mutant copy
166 of *fbxA* (encoding F-box protein A) are mixed with wild-type cells, they become
167 overrepresented as spores rather than stalk cells (Ennis *et al.* 2003). But unlike the facultative
168 cheaters, there is a large group cost; as the frequency of *fbxA* mutants in a group increases, the
169 overall spore production decreases (Gilbert *et al.* 2007). And when alone (*i.e.,* only *fbxA*
170 mutants are present), few to no spores are produced; the loss of *fbxA* results in non-fruiting
171 obligate cheaters (Ennis *et al.* 2000, 2003).
172
173 The cheater genes that have been described thus far in *D. discoideum* share few sequence
174 features or protein domains and are associated with a wide diversity of cellular functions and
175 pathways, which suggests that cheating can be accomplished in many ways (Santorelli *et al.*
176 2008). Some of the mechanistic diversity is already apparent; the described cheater genes

177 cause cheating via different means including altered communications in cell-fate determination
178 (Ennis *et al.* 2000; Thompson *et al.* 2004; Khare and Shaulsky 2010) and changes in cell
179 adhesion (Queller *et al.* 2003). Perhaps that is to be expected given the complexity of
180 multicellular development in *D. discoideum*. In addition to the aggregation and cooperation
181 required of the previously free-living individuals, multicellular development involves the
182 differentiation of cell types (*e.g.,* prestalk, prespore), followed by the maintenance and
183 coordination of each of those cell types fates and their proportions (Loomis 2015). Cheaters can
184 presumably arise that exploit any number of aspects of this developmental process; the study
185 of cheating therefore also has the potential to provide insight into development more broadly.
186
187 To expand our understanding of the genomics underlying cheating and cooperation in *D.*
188 *discoideum,* and especially the understudied obligate cheaters, in this study we return to cell
189 lines experimentally evolved under low-relatedness by Kuzdzal-Fick *et al.* (2011). Starting from
190 a single isolate of the wild type lab strain AX4, Kuzdzal-Fick *et al.* (2011) established 24 replicate
191 lines and maintained low-relatedness by transferring a million spores to new plates 31 times
192 (about 290 generations) (Figure 1A). Each round began with one million spores which were
193 allowed to hatch, proliferate by eating bacteria, and then form fruiting bodies. At each transfer,
194 the low relatedness was reestablished by replating a million thoroughly mixed spores. In this
195 way, cheaters that appeared by mutation would not be with other such cells among the million
196 and would instead be in close proximity to cells to exploit. Under these conditions, cheaters
197 were favored by selection and significantly increased in abundance by contributing more to
198 spores while contributing less to the stalk. As the percentage of cheaters in a line increased,

199 fewer spores were produced overall. And among the evolved lines, the percentage of clonal
200 isolates which had completely lost cooperation and could no longer form fruiting bodies on
201 their own (non-fruiters) rose to as high as 69% (Figure 2). On average, the evolved lines of
202 Kuzdzal-Fick *et al.* (2011) contained approximately 30% non-fruiters and most of them that
203 have been tested were cheaters.
204
205 The work by Kuzdzal-Fick *et al.* (2011) showed that drastically reduced relatedness allowed the
206 spread of cheater mutations that contribute disproportionately to spores and reduced the
207 overall spore production in many lines (Figure 1A). This demonstrates how low relatedness in a
208 natural population would lead to a collapse of multicellularity and with it, the advantages of
209 fruiting body formation and spore dispersal. In this study we use whole-genome sequencing
210 and variant analysis of these experimentally evolved *D. discoideum* cell lines to identify genomic
211 changes that allowed non-fruiting obligate cheaters to evolve over the course of that
212 experiment.
213

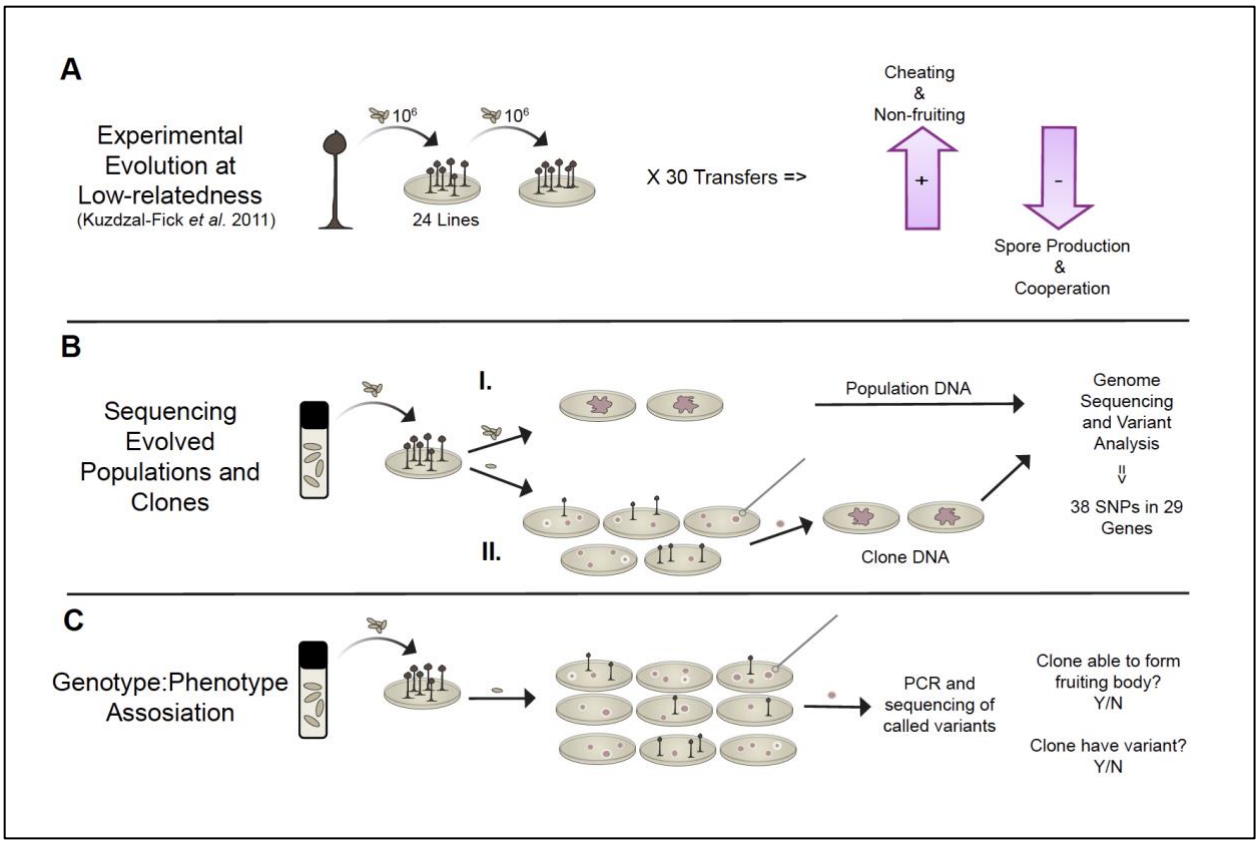
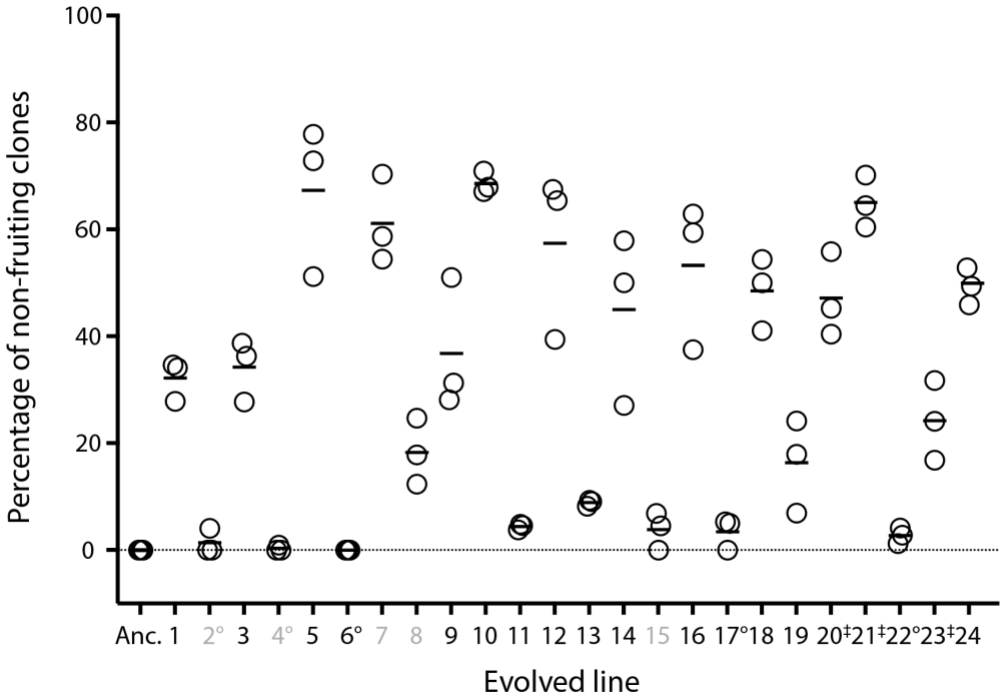


Figure 1. Outline of the experimental workflow. A. Experimental evolution of *D. discoideum* at low-relatedness by Kuzdzal-Fick *et al.* (2011). A clonal isolate of AX4 was used to generate 24 replicate lines. After fruiting body formation, spores were collected, thoroughly mixed and one million spores were replated. Reestablishing low relatedness at each of the 31 passages (about 290 generations) allowed selection to favor mutations that conferred cheating. **B.** Whole-genome sequencing and variant analysis in this study of the 24 evolved cell lines (from A). We grew spores from the final passage of each evolved cell line to generate two samples, (**I.**) the line in bulk, as a population and (**II.**) a non-fruiting clone from each line. For the line in bulk as a population (**B, I.**), we grew cells for genomic DNA extraction. For the non-fruiting clone (**B, II.**), we plated a serial dilution of spores from each population to isolate a non-fruiting clone which we then grew for genomic DNA extraction. We used whole-genome sequencing and variant

226 analysis of populations and clones to identify 38 SNPs impacting 29 genes. **C.** Association of
227 SNPs with the loss of altruism in clones. We again plated the evolved populations (lines) (from
228 A) to allow growth of clonal colonies from individual spores. We evaluated numerous clones
229 from each line by two criteria: 1) whether they were obligate cheaters (non-fruiters) or not
230 (normal fruiting body formation) and 2) whether they carried the previously identified SNP or
231 not, via PCR and Sanger sequencing.
232
233
234 **Materials and Methods**
235 ***Experimental Cell Lines***
236 *Dictyostelium discoideum* cell lines were experimentally evolved under conditions of low-
237 relatedness by Kuzdzal-Fick *et al.* (2011) (Figure 1A). Kuzdzal-Fick *et al.* (2011) froze spores from
238 the evolved lines in KK2 buffer (2.25 g KH_2HPO_4 and 0.67 g K_2HPO_4 per L) with 25% glycerol and
239 stored them at -80°. We thawed spores from the ancestor and from the final passage of each
240 evolved line for genomic DNA extraction. Each evolved line is a population composed of cells
241 and lineages carrying any newly acquired mutations. To capture all of this variation, we
242 generated a whole-genome sequence for each evolved line. We will refer to these sequences as
243 populations. To narrow the focus to variation associated with obligate cheating, we also
244 generated a whole-genome sequence for one non-fruiting clone from each evolved line (Figure
245 1B). We will refer to these sequences as non-fruiter clones. And finally, to identify variation that
246 arose during the experimental evolution we generated a whole-genome sequence of the
247 ancestor.

248



249

250 **Figure 2. Percentage of non-fruiting clones per evolved cell line from Kuzdzal-Fick *et al.***

251 **(2011).** After about 290 generations under low-relatedness, 19 of the 24 lines had evolved to

252 cheat their ancestor (the five lines that did not cheat are labeled in grey). Non-fruiting (obligate

253 cheating), which was not present in the ancestor (Anc.), also increased in frequency to a varying

254 degree among the evolved lines. The percentage of non-fruiting clones for each line (three

255 replicate measurements) is displayed in this scattered dot plot with the mean represented by

256 the horizontal bar. In this study we generated a whole-genome sequence for each evolved line

257 in bulk as a population and for one non-fruiting clone except where indicated ('°' indicates no

258 clone was sequenced and '‡' indicates two clones sequenced). Redrawn from Kuzdzal-Fick *et al.*

259 (2011).

260

261

262 To isolate genomic DNA from each of the 24 evolved lines (Figure 1B, I.), we plated spores onto
263 two SM/5 agar plates (2 g glucose, 2 g BactoPeptone (Oxoid), 2 g yeast extract (Oxoid), 0.2 g
264 $MgCl_2$, 1.9 g KH_2PO_4, 1 g K_2HPO_4 and 15 g agar per liter) with 200 µL of *Klebsiella pneumoniae* in
265 KK2 buffer (OD_{600} 1.5) as food. We plated the spores at two different concentrations (2 X 10^5
266 spores and 4 X 10^5 spores per plate) to allow processing of multiple samples at one time in case
267 growth rates varied between the evolved lines. We incubated the plates at room temperature
268 for approximately 36 h or until log phase. We collected amoebae in log phase from the surface
269 of two plates for each sample and washed them four to five times in chilled KK2 buffer to
270 remove the food bacteria before DNA extraction.

271

272 To isolate genomic DNA from non-fruiting clones (Figure 1B, II.) we first plated serial dilutions of
273 spores to allow for clonal growth from individual spores. We aimed to collect one non-fruiting
274 clone from each evolved line but for five lines (2, 4, 6 and 17 and 22), despite repeated
275 attempts, we were unable to locate any non-fruiting clones. To use the remaining sequencing
276 space, we sequenced a second non-fruiting clone for three haphazardly selected lines (20, 21
277 and 23). After three to five days of growth, we collected cells from the leading edge of each
278 non-fruiting clonal plaque using a sterile loop and plated them on SM/5 agar with *K.*
279 *pneumoniae*. We then plated these cells to grow for DNA extraction using the same protocol
280 described for the evolved lines. To document the non-fruiting morphology, we took
281 photographs of each clonal plaque immediately before collecting cells from the leading edge for
282 expansion (example images are available in Supplemental Material 1). A second photograph

283 was taken one to two days following collection and we continued to monitor the plaques to
284 ensure they never formed fruiting bodies.
285
286
287 ***DNA isolation and sequencing***
288 We isolated genomic DNA from the washed, log-phase cells in the range of 1–2 X 10^8 cells using
289 the Qiagen DNeasy® Blood and Tissue kit (Qiagen). We resuspended the genomic DNA in 10
290 mM Tris-HCl pH 8.5 and stored it at 4° until submission to the McDonnell Genome Institute at
291 Washington University in St. Louis, MO for library preparation and sequencing. Sequence
292 libraries were prepared starting with 0.5 μg of genomic DNA using the KAPA Hyper Library Prep
293 (KAPA Biosystems). We sequenced genomic DNA on the Illumina NovaSeq (150 bp x 2 paired-
294 end) to an estimated depth of 100X and 500X for the clones and populations, respectively.
295
296
297 ***Sequence Alignment***
298 We aligned the Illumina paired-end reads to a single FASTA file containing the reference
299 genomes of both *D. discoideum* AX4 and the food bacterium, *K. pneumoniae* (both downloaded
300 from NCBI in June 2019). We used BWA-MEM (0.7.15) (Li 2013) to index the *D. discoideum* and
301 *K. pneumoniae* concatenated reference genome and to align the paired-end reads for each
302 sample.
303

304 We used an alignment pipeline to run BWA-MEM (Li and Durbin 2009) which converts, sorts
305 and indexes input, intermediate, and output formats using Picard v2.18.1
306 (http://broadinstitute.github.io/picard/), Sambamba v0.6.4 (Tarasov *et al.* 2015) and Samtools
307 v1.3.1 (using HTSLib v1.3.2) (Li *et al.* 2009) ultimately resulting in an aligned, sorted,
308 compressed and indexed CRAM file. We used both Picard (CollectInsertSizeMetrics,
309 CollectAlignmentSummaryMetrics, CollectGcBiasmetrics) and Samtools (flagstat) to evaluate
310 alignment and coverage metrics using BAMs of the initial alignments including both *D.*
311 *discoideum* and *K. pneumoniae* reference genome alignments and again after excluding reads
312 that aligned to *K. pneumoniae*.
313
314
315 ***Variant Calling and Filtration with GATK***
316 We calculated the initial genotype likelihoods using GATK (4.1.2) (McKenna *et al.* 2010)
317 HaplotypeCaller (-ERC GVCF --sample-ploidy 1) for each sample. Next, we ran GATK
318 GenotypeGVCFs (--sample-ploidy 1) to create per-sample genotypes as individual VCF files for
319 the entire *D. discoideum* and *K. pneumoniae* concatenated reference genome. We selected the
320 *D. discoideum* chromosomes (and the unplaced contigs associated with the reference genome)
321 from the VCF files for downstream annotation and filtering using GATK SelectVariants. We used
322 the Ensembl Variant Effect Predictor (VEP 95.3) (McLaren *et al.* 2016) to annotate all variants
323 and add sequence ontology terms (--term SO) using the dicty2.7 assembly of
324 '*Dictyostelium_discoideum*' from Ensembl Protists (release 43). We decomposed complex
325 variants using vt decompose (Tan *et al.* 2015) before adding allele frequencies, merging or

326 filtering. We then processed the decomposed variants with bam-readcount (0.7.4)
327 (https://github.com/genome/bam-readcount) and cyvcf2 (https://github.com/brentp/cyvcf2)
328 to generate allele frequencies. We used VAtools (https://github.com/griffithlab/VAtools, 3.1.0)
329 to add the AF (allele frequency) format fields to each sample including the allele frequencies of
330 each allele as calculated by bam-readcount allele counts. We processed all per-sample VCFs
331 with bgzip and tabix for speed and storage before merging to generate the full per-sample call
332 set VCF using GATK (3.6) CombineVariants (-genotypeMergeOptions UNIQUIFY).
333
334 We performed variant filtration of the GATK VCF using bcftools v1.12 (using HTSLib v1.12) (Li
335 2011) and GATK. We did not consider indels or sites with more than one alternate allele. We
336 are only interested in variation that arose during the course of experimental evolution or, new
337 variation between the ancestor and the evolved lines. To exclude pre-existing variation we first
338 removed sites (from all samples) for which the ancestor was called as a variant (*i.e.,* sites in the
339 ancestor that differed from the reference genome) as well as sites that were left uncalled in the
340 ancestor. To remove ancestral polymorphism we next calculated the major allele frequency
341 (MAF) of all sites in the ancestor BAM file using bam-readcount (0.7.4)
342 (https://github.com/genome/bam-readcount). Using that information, we then removed sites
343 (from all samples) for which in the ancestor, the MAF < 0.90.
344
345 We applied hard filters using GATK VariantFiltration and SelectVariants following the GATK Best
346 Practices standard recommendations (QD > 2, FS < 60, SOR < 3, MQ > 40, MQRankSum > -12.5
347 and ReadPosRankSum > -8). Next we applied custom filters using bcftools view. The first custom

348 filter was the removal of sites missing too much data, which we defined as sites left uncalled in
349 more than 10 samples. For each of the 24 evolved lines we sequenced the whole line (as a
350 population), and for most of the lines we also sequenced one non-fruiting clone, for a total of
351 two samples per line (or a total of three samples for lines 20, 21 and 23 for which we
352 sequenced two clones). There is a low likelihood that the same SNP will occur by chance in
353 more than one evolved line. But because we sequenced two or three samples for each line (the
354 population and individual clones) the maximum number of times that a SNP is likely to appear
355 in our data is twice (or three times for lines 20, 21 and 23). For this reason, we applied a
356 maximum alternate allele count (AC) of three. Next, we applied a minimum Phred-based quality
357 score (QUAL) of 200 to remove low-quality sites and we removed sites with more than 1.5
358 times the average approximate read depth (DP) to reduce false positives. Last, we manually
359 reviewed this final set of SNPs using the Integrative Genomics Viewer (IGV) to further reduce
360 the number of false positives and misclassifications (Robinson *et al.* 2011, 2017).
361
362
363 ***Variant Calling and Filtration with Freebayes***
364 We also performed variant calling on all samples (joint calling) using Freebayes v1.3.1-dirty
365 (Garrison and Marth 2012). We used the default parameters but for the following exceptions:
366 sample ploidy of one, pooled continuous mode, minimum base quality 10, minimum mapping
367 quality 10 and we only retained the best of six alleles. We streamed variant calls directly
368 through the vcflib (Garrison *et al.* 2021) vcffilter, to remove variants with a quality score below

369 20. We decomposed complex variants into their constituent SNPs and indels using vcflib
370 vcfallelicprimitives followed by normalization with vt (Tan *et al.* 2015).
371
372 We performed variant filtration of the normalized Freebayes VCF using bcftools v1.12 (using
373 HTSLib v1.12) (Li 2011) following a similar process to that used for filtering the GATK VCF.
374 Although the two callers calculate and output some different metrics for quality assessment,
375 we generated filters resembling those applied to the GATK VCF as much as possible. First, we
376 excluded indels and sites with more than one alternate allele, followed by the removal of
377 background variation as described for the GATK VCF (ancestral sites with an MAF < 0.90 or sites
378 in the ancestor that were either called as a variant or left uncalled). Next we applied hard filters
379 to remove calls affected by mapping quality or strand bias (MQM > 40, SAF > 0 and SAR > 0, SAP
380 > 0.5 and SRP > 0.5, RPR > 1 and RPL > 1). We then applied the same set of custom filters to the
381 Freebayes VCF as described for the GATK VCF including the removal of sites with more than 10
382 missing samples, sites with an alternate allele count (AC) greater than three and sites with a
383 Phred-based quality score (QUAL) below 200. Finally, based on the approximate read depth
384 (DP) in this remaining set of variants we removed sites with more than 1.5 times the average to
385 reduce false positives. To further reduce the number of false positives and misclassifications we
386 manually reviewed this final set of SNPs using the Integrative Genomics Viewer (IGV) (Robinson
387 *et al.* 2011, 2017).
388
389 ***The Intersection of GATK and Freebayes VCFs and Variant Read Support***

390 We viewed the two separately generated and filtered VCF files (from GATK and Freebayes) side-
391 by-side and removed four sites that were not present in both files. The SNPs we removed
392 included one SNP that had only been initially called by one of the callers and three SNPs that
393 although initially called by both callers, only survived the filtration in one of the separate
394 filtering pipelines. Going forward we worked with these cross-validated SNPs in the annotated
395 GATK generated VCF.
396
397 As one final verification of variant support and quality we confirmed that each variant that had
398 been called in a non-fruiting clone could also be detected in its origin population (or the line
399 from which the clone was isolated). To do this, we generated read counts for each SNP across
400 all samples using bam-readcount v0.7.4 (https://github.com/genome/bam-readcount). We
401 removed any SNP for which the origin population did not support the same alternate allele with
402 at least five reads, which we expect would only occur for false positives or for SNPs that did not
403 provide a selective advantage. Also, because we expect that most true variants will be unique
404 to one replicate line, for each SNP we viewed the number of reads supporting the same
405 alternate allele across all samples. This led to the removal of one SNP that shared low level
406 support for the same allele across multiple lines and also the rejection of one sample from a
407 variant call (for details see *Variant Read Support* in the Supplemental Methods).
408
409
410 ***Functional Annotation Clustering of Genes with SNPs***

411 We performed functional annotation enrichment analysis with the final set of 29 genes
412 containing SNPs using the online tool, DAVID (Database for Annotation, Visualization and
413 Integrated Discovery (https://david.ncifcrf.gov/; Huang *et al.* 2009a, 2009b). We determined
414 enriched terms using a Benjamini-corrected p-value (< 0.05) (Benjamini and Hochberg 1995).
415
416
417 ***Structural Variant (SV) Calling and Filtration with Delly***
418 We called structural variants (SVs) with Delly v0.8.3 (Rausch *et al.* 2012) following the Germline
419 SV Calling workflow. We first did per-sample SV calling, providing the indexed, sorted and
420 duplicate-marked BAM files and the indexed *D. discoideum* and *K. pneumoniae* concatenated
421 reference genome. We then merged the SV sites to a single list and called genotypes across all
422 samples. We merged all samples genotypes using bcftools v1.12 (using HTSLib v1.12) (Li 2011).
423 We applied the Delly filter (-f germline) and then removed low quality variants that were not
424 flagged as "PASS" in the VCF filter field (PE > 3 [or PE > 5 for translocations] and QUAL >=20).
425 We annotated the SVs using the Ensembl Variant Effect Predictor (VEP 95.3) as described for
426 the GATK VCF. We only retained simple, intrachromosomal SVs. As described for the SNPs, the
427 same SV is unlikely to occur in more than one evolved line. But because we sequenced two or
428 three samples for each line (the population and individual clones) the maximum number of
429 times that a SNP is likely to appear in our data is twice (or three times for lines 20, 21 and 23).
430 For this reason, we discarded SVs that were called in more than three samples. Last, as
431 described for the SNPs, we manually reviewed this final set of SVs in IGV (Robinson *et al.* 2011).
432

433

434 ***Association of Variants with the Loss of Altruism***

435 To investigate for a potential correlation between variants and the complete loss of
436 cooperation we returned to some of the evolved lines to isolate additional clones for targeted
437 re-sequencing (Figure 1C). Each evolved line is a population comprised of both cooperators and
438 evolved cheaters. The complete loss of cooperation is therefore only revealed when spores are
439 plated clonally and can be identified by an inability to form a fruiting body (non-fruiting clones).
440 Taking advantage of this population variation (Figure 2) we scored multiple additional clones for
441 1) whether or not it was able to form a fruiting body in isolation and 2) the presence or absence
442 of the called variant(s) in the line.

443

444 For this analysis we again plated the evolved cell lines clonally (as described in the section,
445 *Experimental Cell Lines*) to allow for growth from individual spores. After three to five days of
446 growth we screened clones via PCR and Sanger sequencing. As previously described (in the
447 section, *Experimental Cell Lines*), we took photographs of the screened clones to document the
448 presence or absence of fruiting body formation (example images are available in Supplemental
449 Material 1).

450

451 We generated genomic DNA for PCR genotyping clones using one of two methods. For some of
452 the clones we first collected cells from the leading edge of the plaque, grew them to larger
453 numbers and carried out a formal DNA extraction with the DNeasy® Blood and Tissue kit
454 (Qiagen). To increase throughput, for most clones we directly lysed cells from the leading edge

455 of the plaque and used the lysate for what we called a "plaque PCR." Based on a protocol
456 described by Charette and Cosson (2004), the plaque PCR included two steps. First, we used a
457 sterile pipette tip to collect a small number of cells from the leading edge of a plaque and
458 placed it in a tube containing 20 µl lysis buffer (10 mM Tris, pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$,
459 0.45% Nonidet® P-40 (NP40) and 0.45% Tween® 20) with PK (1 µL of 20 µg/µL of PK for every 25
460 µL of lysis buffer). Next, we incubated the cells in lysis buffer for 1 min at 95° to inactivate the
461 PK after which we used the cell lysate for PCR or stored them at -20°.
462
463 We designed primers to PCR-amplify the region spanning each called variant of interest
464 (Supplemental Table S1). For the PCR we used either 1 µl of cell lysate or approximately 10 ng
465 of DNA for the formal DNA isolations and the following reaction components: $MgCl_2$ (25 mM) 1
466 µl, dNTPs (10 mM each) 0.5 µl, 5 pM of each primer, 5X GoTaq® Flexi Buffer 5 µl, GoTaq® DNA
467 Polymerase 0.2 µl and H_20 10.3 µl for a 20 µl reaction. We used the following PCR protocol
468 (adjusting the annealing temperature as needed, according to primer pairing): 95°, 2:00; 95°,
469 0:15; 50°, 0:15; 60°, 3:00; repeat steps 2–4 34X, 60°, 5:00; 4° hold. We submitted PCR products
470 and primers (the same primers used for amplification) to Genewiz (South Plainfield, NJ) for
471 purification and Sanger sequencing. We trimmed the returned sequences with 4Peaks
472 (https://nucleobytes.com/4peaks/) for alignment using Seaview
473 (http://doua.prabi.fr/software/seaview).
474
475 We screened a varying number of clones according to the availability of the fruiting and non-
476 fruiting clones from each line. For each line included in this screen we tallied the number of

477 clones that did and did not carry the called mutation(s) and whether or not the clone was able
478 to form a fruiting body when plated clonally. We tested for significance using Fisher's Exact Test
479 with a 95% confidence interval using GraphPad Prism (version 9.3 for MacOS).
480
481
482 ***Predicted Protein Structure***
483 To explore proteins of interest, we first downloaded the amino acid sequences from DictyBase
484 (Fey *et al.* 2009) and submitted them to Protter (Omasits *et al.* 2014) for visualization. We
485 accessed three-dimensional structural predictions generated by AlphaFold
486 (https://alphafold.ebi.ac.uk/) (Jumper *et al.* 2021; Varadi *et al.* 2022).
487
488
489 ***Data Availability Statement***
490 Raw reads in CRAM format will be stored in the Sequence Read Archive (SRA) [link/citation]
491 (accession numbers/DOIs). Variant calls that survived filtration from each of the three variant
492 callers (GATK, Freebayes and Delly) are available in Supplemental Material 2 in VCF format. All
493 cell lines used in this study are in the lab of the senior authors.
494
495
496 **Results**
497 ***Raw Data Generation and Alignment***

498 We obtained over 5 trillion reads across all samples, with a high rate of alignment (94.7%) to
499 the concatenated *D. discoideum* and *K. pneumoniae* reference genomes. All reads that aligned
500 to *K. pneumoniae* (food bacterium) were excluded from further analysis. The remaining 2.7
501 trillion reads aligned to *D. discoideum* with an average rate of 89.3%. Read alignment was, on
502 average, equally successful for clones and evolved lines (89.3% and 89.4%, respectively). The
503 average mapped read depth of all annotated genes normalized by gene length (gene
504 annotations downloaded from NCBI on 10-25-19, Supplemental Material 3) is 54X and 195X for
505 clones and evolved lines, respectively (Table 1). The average mapped read depth of normalized
506 intergenic regions is very similar with averages of 53X and 180X for clones and evolved lines,
507 respectively.
508
509
510 ***Resulting SNPs are Selection Driven***
511 We called SNPs across all samples (joint calling) resulting in 53,220 unfiltered variants with
512 GATK and 32,775 with Freebayes (before normalization). After independently filtering each call
513 set the GATK VCF contained 306 SNPs and the Freebayes VCF contained 230 SNPs (the number
514 of variants removed by each filter is available in Supplemental Material 4). We next manually
515 reviewed all remaining SNPs in IGV resulting in approximately 75 sites in each call set. Finally,
516 we removed any sites that did not co-occur in both call sets and/or sites lacking read support
517 (as described in ***The Intersection of GATK and Freebayes VCFs and Variant Read Support***). The
518 final set of SNPs contained 38 biallelic SNPs associated with 29 different genes (Table 2). Each
519 gene in this list has only one SNP except for the gene, *grlG* (DDB_G0272244), which has 10

520 unique SNPs (discussed in detail in the following sections). All SNPs are unique to one evolved
521 line with the exception of one in the unannotated gene, DDB_G0276529, which was called in
522 two lines (7 and 16). Most evolved lines have between one and three SNPs except for lines 7
523 and 21, which have five and eight SNPs each, respectively (Table 1).
524
525 Most SNPs (31 of 38) are in coding regions of the genome and all but one of these result in an
526 introduced stop codon or amino acid substitution (Table 2). Among the 31 SNPs in coding
527 sequence, 22 are missense variants and according to the Variant Effect Predictor annotations
528 are predicted to have a moderate impact, eight others introduce a premature stop codon with
529 a high predicted impact and the one synonymous variant has a low predicted impact. Among
530 the seven SNPs in non-coding sequence, five are upstream and two are downstream variants
531 that the Variant Effect Predictor annotated as "modifier," meaning they are either difficult to
532 predict or that there is no evidence of an impact. This distribution of large effect SNPs strongly
533 supports that they have increased in abundance due to selection, rather than drift.
534
535
536 ***Genes with SNPs***
537 The 29 genes with SNPs (Table 2) are distributed throughout the *D. discoideum* genome with
538 between one and eight SNPs on each of the six chromosomes. The average GC content for the
539 set of 29 genes (excluding introns) is 28.4% which is close to the genome-wide average of 27%
540 for protein-coding genes. Each of the genes in our list has one SNP except for the G protein-
541 coupled receptor (GPCR), *grlG*, which has 10 unique SNPs. *GrlG* is one of 17 glutamate receptor-

542 like ('grl') proteins in the *D. discoideum* genome (*grlA*–H and *grlJ*–R) (Eichinger *et al.* 2005) but
543 the function of *grlG* is not yet known. And like the majority of glutamate receptor-like proteins
544 in *D. discoideum*, we do not know the ligand(s) that it binds nor the downstream effector(s).
545 Among the remaining 28 genes (each with one unique SNP) a few have been well described
546 (*e.g., sgmA*, *pks39* and *arkA*) but the majority (17) are hypothetical proteins still largely lacking
547 annotation.
548
549 Using annotations that are available for our 29 genes with SNPs, we carried out functional
550 annotation clustering with DAVID and identified two significantly enriched annotation clusters.
551 The first cluster of five annotations contained UniProt Keywords (Bateman *et al.* 2021) related
552 to zinc and metal binding (group enrichment score of 1.45) including *arkA*, DDB_G0272484,
553 *rnf160*, DDB_G0269332 and *sgmA*. The second cluster of 10 annotations contained terms
554 related to transmembrane and membrane annotations (group enrichment score of 1.11). This
555 second cluster contained *arkA* and DDB_G0269332 from the first cluster as well as *grlG*, *pks39*,
556 DDB_G0276291, DDB_G0278531, DDB_G0278575, DDB_G0281923, DDB_G0290523 and
557 *tmem144A*. After applying the Benjamini (Benjamini and Hochberg 1995) correction, none of
558 the individual terms are significantly enriched (Supplemental Material 5). The large number of
559 unannotated genes implicated here does hinder interpretation, but it is not unusual given that
560 at the time of writing this manuscript, approximately 40% of protein-coding genes in the *D.*
561 *discoideum* genome still lack annotation (Fey *et al.* 2009).
562
563

564 ***Structural Variants (SVs) Provide Further Support for grlG***

565 We called structural variants (SVs) across all samples (joint calling) using Delly resulting in an
566 unfiltered set of 10,139 SVs. The Delly filter reduced the number to 131 and after all quality
567 filtration we have a set of 12 SVs (Supplemental Table S2). Among the 12 SVs are six deletions,
568 five inversions and one duplication. Nine of the 24 evolved lines carry 1 or more of these SVs,
569 two of which (lines 1 and 14) did not have any called SNPs. Each SV is unique to one evolved
570 line with the exception of one deletion (NC_007088.5:1742215–1742506) in *grlG* which was
571 called in lines 17 and 18. However, because we cannot rule out potential contamination of the
572 population sample of line 18 we only report the variant in line 17 (more detail is available in
573 *Variant Read Support* in the Supplemental Methods). The deletion was verified by PCR in line 17
574 (as part of the analysis described in the section, *Association of Variants with the Loss of*
575 *Altruism*). The SVs range considerably in length (from 183 bp to 49 Kb) potentially impacting as
576 many as 21 different genes. Strikingly, five of the 12 SVs impact *grlG*, the same gene that we
577 already identified as carrying 10 SNPs. None of the 28 other genes with SNPs were impacted by
578 any of the SVs. Given the clear signals of parallel evolution identified for *grlG*, we chose not to
579 pursue the other SVs any further. We will only discuss the SVs that impact *grlG* going forward;
580 the full set of 29 SVs that passed filtration are available in VCF format in Supplemental Material
581 2.

582

583

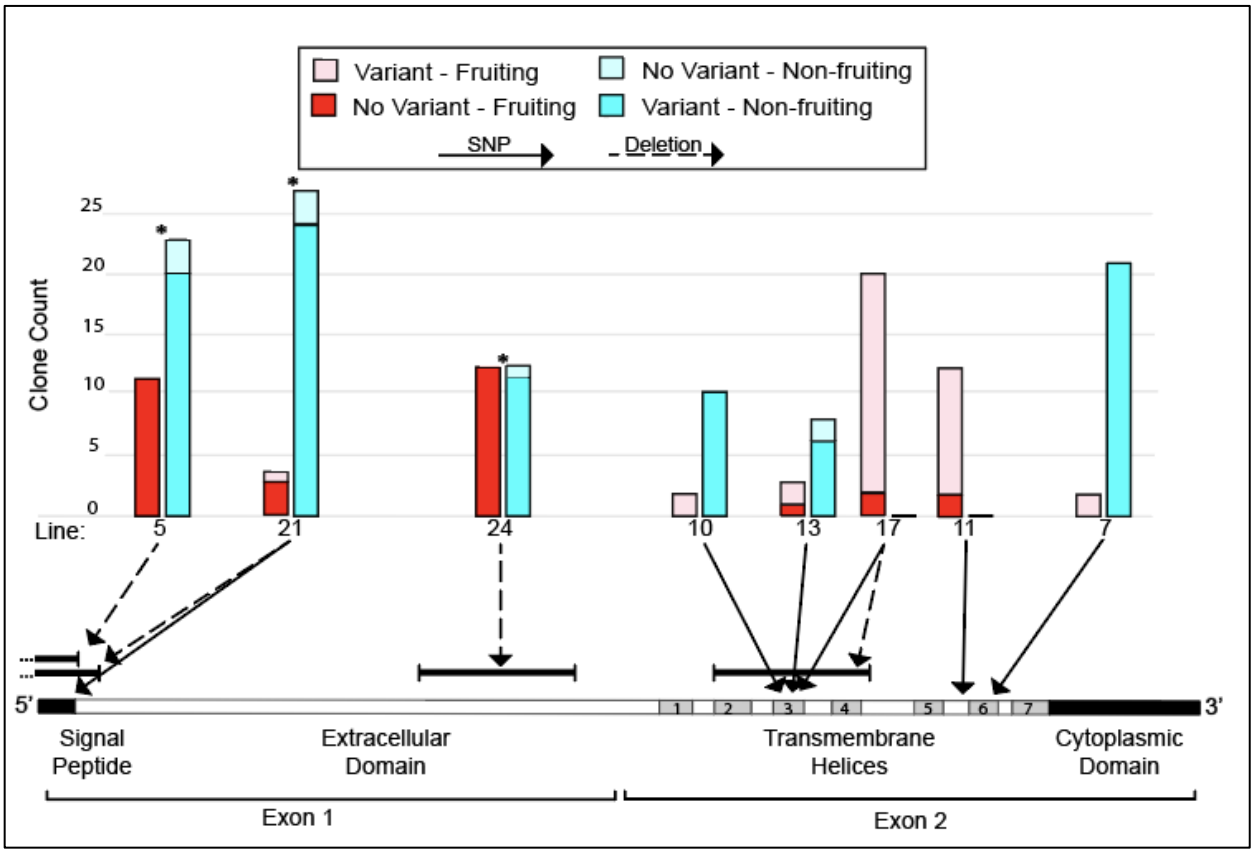
584 ***Some Variants in grlG are Associated with the Loss of Cooperation***

585 Given the high level of parallelism identified in *grlG* (one or more variants in more than half of
586 the evolved lines) we returned to the evolved lines to screen additional fruiting and non-fruiting
587 clones for a correlation between variants in *grlG* and the loss of cooperation. The variants are
588 located throughout the length of the gene but a pattern emerged in which only those variants
589 located in the 5' half of *grlG* (on exon 1) are associated with the loss of cooperation (*i.e.,* non-
590 fruiting clones) and variants in the 3' half (on exon 2) are not (Figure 3 and Supplemental Table
591 S3). Because *grlG* has two exons of roughly equal length (383 and 390 amino acids for exon 1
592 and exon2, respectively) and for simplicity, we will refer to the two halves as the 5' and 3'
593 regions.
594
595 We prioritized screening of variants that we had identified in evolved lines with a moderate
596 percentage of non-fruiting clones (Figure 2) so that both fruiting and non-fruiting clones could
597 be included. We also prioritized variants with moderate allele frequencies in their evolved
598 populations. As such, the analysis included ten of the 15 variants identified across the length of
599 *grlG*. The 10 variants were called in 8 different lines (two unique variants were called in lines 21
600 and 17), three of which have variants in the 5' region (lines 5, 21 and 24) and five with variants
601 in the 3' region (7, 10, 11, 13 and 17). From the three lines with variants in the 5' region, we
602 screened 88 clones (62 non-fruiters and 27 fruiters) for the presence of four called variants. We
603 found that 88.7% (55 of 62) of the non-fruiter clones carried the variant compared to only 3.7%
604 (1 of 27) of the fruiting clones. The association between the presence of variants in the 5'
605 region of *grlG* and the clonal phenotype is significant in each evolved line we screened (Fisher's
606 Exact $p < 0.0001$, $p = 0.0164$ and $p < 0.0001$ for lines 5, 21 and 24, respectively). In the 3' region,

607 from the five lines with variants, we screened 78 clones (39 non-fruiters and 39 fruiters) for the
608 presence of six called variants. Unlike the association in the 5' region, we found the number of
609 clones carrying a variant in the 3' region was roughly equal between non-fruiters and fruiters
610 (94.8% [37 of 39] and 87.2% [34 of 39], respectively). There is no association between the
611 presence of variants in the 3' region of *grlG* and the clonal phenotype in any of the evolved
612 lines we screened (Fisher's Exact $p > 0.05$).
613
614 According to our results, variants in the 5' region of *grlG* are significantly associated with the
615 non-fruiting phenotype but not all clones follow that trend (Figure 3). In all three evolved lines
616 that we screened in the 5' region, we found a few clones that did not have a variant, but they
617 were still unable to form fruiting bodies in isolation. These clones could have lost cooperation
618 because of a different, undetected variant (anywhere in the genome). However, a clone that is
619 still able to cooperate and form a fruiting body in isolation, despite the apparent loss of *grlG,* is
620 more difficult to align with our hypothesis that high impact variants in *grlG* cause non-fruiting.
621 We only found one such clone in the 5' region. One clone from line 21 was able to form a
622 fruiting body despite carrying a SNP that introduces a premature stop codon in *grlG* after only
623 23 amino acids. None of the other 26 fruiting clones that we screened in the 5' region carried a
624 variant in *grlG*. This is strikingly different from the results in the 3' region, where almost all of
625 the fruiting clones that we screened carried a variant in *grlG,* though these were concentrated
626 in lines 11 and 17 and rare in others. Thus, variants in the 5’ region of *grlG* usually result in the
627 loss of cooperation but variants in the 3’ region do not. For a subset of the clones that we
628 screened for *grlG* variants we also screened for the presence of one or more SNPs that we

629 called in other gene(s). These additional screens were largely uninformative and are therefore

630 only described in the supplemental materials (Additional File 1 and Supplemental Table S4).



631

632 **Figure 3. Location of variants in *grlG* and correlation with the loss of cooperation.** Schematic

633 representation (drawn to scale) of the coding sequence of *grlG*. The major features of a mature

634 *grlG* protein are indicated below the schematic (see also, Figure 5)**.** Ten of the 15 called variants

635 (labeled above the schematic) were included in a targeted resequencing of additional clones

636 from each respective evolved line to investigate the association between variants in *grlG* and

637 the loss of cooperation (clonal inability to form a fruiting body). For each evolved line that we

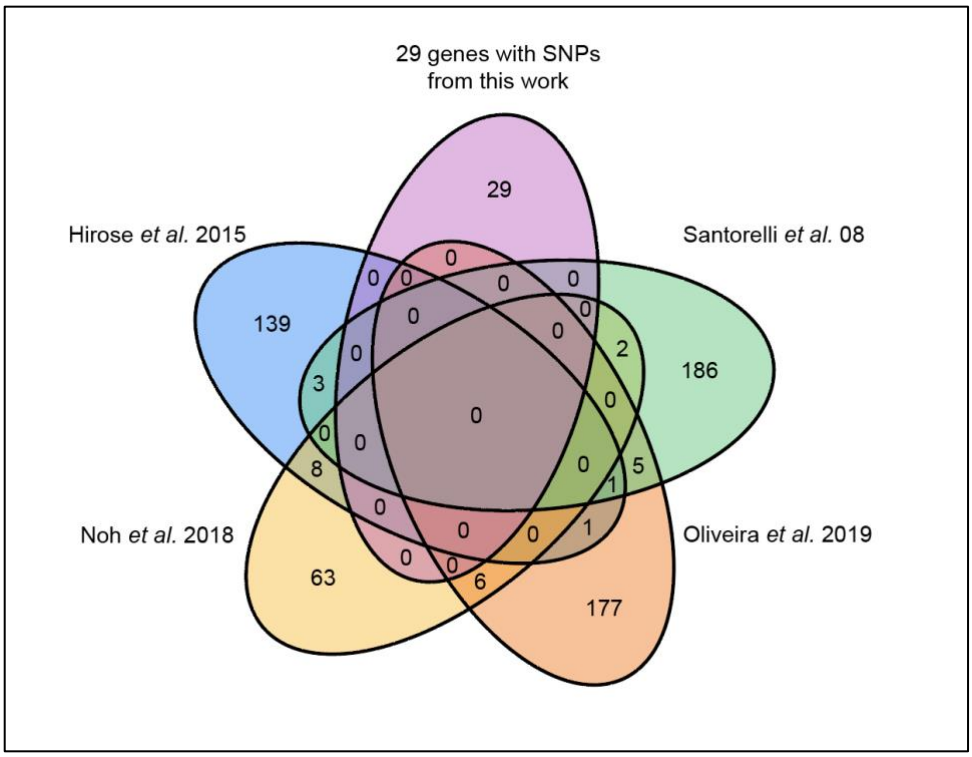
638 screened clones from, a histogram illustrates the results with side-by-side bars showing the

639 number of clones that could (left bar) and could not (right bar) form a fruiting body in isolation

640 and the number of each phenotype (fruiting or non-fruiting) that did or did not carry the called
641 variant. Significant associations are indicated by an asterisk (*).
642
643
644 ***The Genes Discovered in This Study Have Not Previously Been Implicated in Cheating***
645 To look for a relationship between the 29 genes identified herein and the loss of altruism, we
646 turned to the literature to see if any of them have been previously implicated in cheating in *D.*
647 *discoideum*. Because we are looking for obligate cheater genes, most notably missing from our
648 list is *fbxA,* the only previously known obligate cheater gene (Ennis *et al.* 2000). It is also
649 interesting that our list does not contain any of the characterized facultative cheater genes such
650 as *chtB, chtC*, *dimA* or *csA* (Queller *et al.* 2003; Foster *et al.* 2004; Khare and Shaulsky 2010;
651 Santorelli *et al.* 2013). Aside from that handful of genes that first come to mind, the remaining
652 majority of genes with potential involvement in cheating in *D. discoideum*, have not yet
653 received attention at the individual level. We were particularly interested in the many loci
654 recovered in the large restriction enzyme-mediated insertional (REMI) mutant library screening
655 from Santorelli *et al.* (2008). In that study, a set of 167 mutants (corresponding to 197 loci)
656 were identified that increased under low-relatedness selection. Based on testing a subset of
657 those mutants, the authors estimate that approximately 75% of the mutants are cheaters, but
658 so far, only *chtB* and *chtC* have been described. Although Santorelli *et al.* (2008) focused on
659 facultative cheater genes (by requiring one round of solo fruiting body formation) they did
660 recover *fbxA* and a few other mutants were found with morphological abnormalities indicating
661 there may be a others that are not facultative. The full set of 197 loci identified by Santorelli *et*

662 *al.* (2008) is the first of four gene lists we chose for cross-comparison. The other gene sets are
663 each comprised of genes found to be differentially expressed in chimeric mixtures identified at
664 various stages of multicellular development and thus may contain genes involved in conflict and
665 cheating. Two are comprised of genes identified at tight aggregate stage (79 genes) or the slug
666 stage (190 genes) (Noh *et al.* 2018 and Oliveira *et al.* 2019, respectively). The last gene set is
667 composed of 152 genes that were also identified during the establishment of multicellularity (at
668 either 8 or 12 hours), but in this case, the strains had been engineered to be mutually
669 incompatible (*tgrB1*/*tgrC1* gene replacement strains) which may increase the potential to
670 identify genes involved in conflict (Hirose *et al.* 2015). Unlike the genes from Santorelli et al.
671 (2008) these last three gene sets have not been directly implicated in cheating. But given their
672 differential expression in chimeras, a time at which cheating or the resistance to cheating
673 would be adaptive (Noh *et al.* 2018), they are likely to include genes involved in conflict and
674 cheating. In support of this reasoning, the 79 genes from Noh *et al.* (2018) show signatures of
675 rapid molecular evolution, consistent with conflict driven escalating arms races (Queller and
676 Strassmann 2018; Scott and Queller 2019). For simplicity, we will refer to these four sets of
677 genes as candidate cheater genes.
678
679 Altogether, the four sets of candidate cheater genes from the literature comprise a list of 591
680 unique genes (Supplemental Table S4). None of the 29 genes identified in this study were found
681 on that list and there is little overlap among the four candidate cheater gene sets (Figure 4).
682 This in line with previous reports of limited overlap among sets of social genes (kin
683 discrimination, cooperation and cheating) (Oliveira *et al.* 2019; Noh *et al.* 2020). Only 26 of the

591 genes are present on more than one list. No genes are present on all four lists and only one gene is present on three of the four; the pre-spore specific gene D (*pspD*). Because the genes from Santorelli *et al.* (2008) were identified as part of a screen for cheater genes specifically, they are of particular interest. Eleven genes from Santorelli *et al.* (2008) were found on one or more of the other gene lists and three of them were among the subset of mutants directly tested for cheating in that study. One of these tested mutants (LAS92) had an insertion in a gene that is also on the list from Hirose *et al.* (2015) and it was confirmed to be a cheater (DDB_G0283373). Two other tested mutants had intergenic insertions near genes that are also on the list from Oliveira *et al.* (2019); one of which was confirmed to be a cheater (LAS6) (DDB_G0276335) and the other was not (LAS60) (DDB_G0285843). However, these last two mutants from Santorelli *et al.* (2008) had intergenic insertions and the actual gene(s) impacted was not precisely determined.


29 genes with SNPs
from this work
29
Hirose et al. 2015
139
Santorelli et al. 08
186
Noh et al. 2018
63
Oliveira et al. 2019
177
0
0
0

710

711 **Discussion**

712 In this study we identified SNPs in 29 genes that arose across 24 replicate lines of *D. discoideum*
713 experimentally evolved by Kuzdzal-Fick *et al.* (2011) under conditions that mean low-
714 relatedness for any new mutations. In that earlier study, over the course of about 290
715 generations at low-relatedness, 19 of the 24 replicate lines saw a significant rise in cheating,
716 defined as a mutant contributing proportionally more to spore cells than stalk cells. When
717 isolates from these lines were plated clonally, an average of 31% had completely lost altruism
718 and were unable to form fruiting bodies on their own. Because each line is a mixed population,
719 the mutations that conferred the obligate cheating advantage were able to spread by forming
720 chimeric fruiting bodies with cooperators lacking the mutation. Here we used whole-genome
721 sequencing and variant analyses of those previously evolved cell lines to identify which genes
722 had changed and might be responsible for the non-fruiting, obligate cheating phenotype. While
723 there are many genes known to cause facultative cheating, before this study only one gene
724 (*fbxA*) had been recognized to cause obligate cheating. We identified at least one SNP in 17 of
725 the 24 experimentally evolved lines for a total of 38 SNPs. Most of the SNPs (31) are in coding
726 regions of the genome, and all but one results in either an introduced stop codon or other
727 amino acid substitution suggesting selective pressure acting on these loci under the
728 experimental conditions in Kuzdzal-Fick *et al.* (2011). The majority of the genes with SNPs are
729 not well characterized and many lack any annotation which impacted our ability to assess any
730 uniting similarities among them. Except for one of these genes, *grlG*, with 10 unique SNPs
731 across multiple lines, the other 28 genes only have one SNP each. *GrlG* also stands out because

732 we found it also harbors five structural variants, whereas none of the other genes with SNPs
733 had any. Our results provide strong evidence of parallel evolution of *grlG* and suggests the
734 involvement of this G protein-coupled receptor (GPCR) in cooperation and multicellular
735 development in *D. discoideum*.
736
737
738 ***Confidence in Variant Calls and Detection***
739 The *D. discoideum* genome has features that may have impacted our ability to detect variants.
740 The AT content of the *D. discoideum* genome is very high, greater than 77% (Eichinger *et al.*
741 2005). The high AT content contributes to the high density of simple sequence repeats (DNA
742 tracts of 1–6 bp tandemly repeated a varying number of times) (Tian *et al.* 2011) which
743 represent 14.3% of the genome (Srivastava *et al.* 2019). Moreover, in *D. discoideum*, these
744 repeats are not confined to non-coding regions; they are also found in over 16.3% of protein
745 coding genes (Eichinger *et al.* 2005). Simple sequence repeats are recognized as mutational
746 hotspots across a number of organisms and are particularly prone to insertions and deletions
747 (indels) (Tóth *et al.* 2000; Ellegren 2004). In *D. discoideum*, both indels and SNPs have been
748 shown to occur at a higher frequency in low-complexity, repetitive regions of the genome
749 (Srivastava *et al.* 2019). The precision of variant calling is also decreased in the same sorts of
750 low-complexity and repeat regions, compounding the challenges of variant calling in *D.*
751 *discoideum*. This decreased precision is particularly true for detection of structural variants,
752 which are inherently more challenging to identify, particularly with short-read sequence data
753 (Cameron *et al.* 2019; Kosugi *et al.* 2019; Mahmoud *et al.* 2019). Despite these challenges, we

754 chose not to exclude low-complexity or non-coding regions of the genome, which could lead to
755 undiscovered variation outright. Instead, we dealt with the challenging features of the *D.*
756 *discoideum* genome by carefully exploring the effects of different settings for variant calling and
757 filtering to determine a stringent and robust pipeline. We separately called SNPs with two
758 independent callers. We applied our carefully chosen quality filters to the two call sets
759 independently, followed by manual inspection of all variants. We then conservatively only
760 pursued the SNPs still present in both call sets. Finally, we used the BAM read counts across all
761 samples to rule out potential contamination or other false positives. Together, despite the
762 challenging features of the *D. discoideum* genome, we are confident that we obtained a high
763 quality and strongly supported set of variants.
764
765 Our stringent pipeline decreased the likelihood of obtaining false positives and ensured that
766 variants called at the population level tend to be those most strongly selected. The trade-off of
767 such a stringent approach is a decreased sensitivity to detect low frequency variants in the
768 population samples. Given the relatively short amount of time the experimental lines were
769 allowed to evolve (Kuzdzal-Fick *et al.* 2011), not all variants would have had the time to expand
770 throughout the population of that line. Therefore, variants that arose later in the experiment or
771 those not facing strong selection, may have been difficult for us to detect at the population
772 level. Although we cannot rule out the possibility that some were left undetected, we did
773 detect some low frequency variants by sequencing individual non-fruiting clones from the lines.
774
775

776 ***Parallel Evolution and the Involvement of grlG in Cooperation***

777 The concentration of highly supported and unique variants in *grlG* across 14 of the 24 replicate
778 lines is strong evidence of parallel evolution at the gene level (Tenaillon *et al.* 2012; Barrick and
779 Lenski 2013; Van den Bergh *et al.* 2018). None of these mutations were present in the ancestral
780 line, indicating that they occurred during the course of experimental evolution (Kuzdzal-Fick *et*
781 *al.* 2011). At the nucleotide level there is no indication of a mutational hotspot because the
782 variants are located throughout the length of the gene. Moreover, the variants in *grlG* are
783 comprised of several types of mutations including 10 non-synonymous SNPs that result in
784 either an introduced stop codon (four) or a missense mutation (six), as well as four deletions
785 and one inversion, all of which impact coding sequence. The number of large-effect mutations
786 (half of which result in protein truncation) and the complete lack of synonymous mutations in
787 *grlG* suggests that the loss of *grlG* was adaptive under the experimental conditions. Further
788 support of selection acting on *grlG* is the high population level frequency that several of these
789 new mutations reached in their respective lines. On average, the 10 unique SNPs in *grlG* are
790 supported by 61% of the total reads in their respective populations compared to an average of
791 only 41% support for the 28 remaining SNPs in other genes. These data suggest that *grlG*
792 variants increased in abundance due to selection, rather than drift. In addition, by screening
793 multiple additional clones from the experimental lines, we found that variants in the 5' region
794 of *grlG* are significantly associated with the loss of cooperation (non-fruiting) while those in the
795 3' region are not. However, very few of the clones screened in the 3' region carried a variant in
796 *grlG* (9%) and this may have limited our power to detect an association. We were unable to find
797 any non-fruiting clones for evolved lines 11 and 17 but the large proportion of normally fruiting

798 clones that carry a *grlG* variant certainly suggests a lack of an association in those lines. To help
799 us make sense of these results, we next explore the limited amount of currently available
800 information about *grlG* and view our findings in light of what is known more broadly about the
801 large and highly conserved family of proteins, the G protein-coupled receptors (GPCRs).
802
803 The G protein-coupled receptors (GPCRs) in eukaryotes are the largest class of receptors of
804 extracellular stimuli. They generally activate intracellular responses via heterotrimeric G-
805 proteins. In humans there are over 800 GPCR genes making them the most abundant gene
806 family in the human genome (Lagerström and Schiöth 2008). They are involved in a huge
807 diversity of biological processes including signaling of hormones, metabolites and
808 neurotransmitters to cell migration and development (Rosenbaum *et al.* 2009). The GPCRs are
809 responsible for approximately 66 known single-gene human diseases, and they are the target of
810 about one third of FDA-approved drugs (Hauser *et al.* 2017; Schöneberg and Liebscher 2021).
811 Thus, identifying the ligands and the roles of yet uncharacterized GPCRs (or, orphan GPCRs) is of
812 great interest broadly and the GPCRs in *Dictyostelium* are no exception. *Dictyostelium*
813 *discoideum* is an established model system for studying numerous processes such as chemotaxis
814 and other cell-motility linked processes such as phagocytosis and development (Bozzaro 2019),
815 all of which involve GPCRs.
816
817 The *D. discoideum* genome has a surprisingly large and diverse repertoire of more than 55
818 GPCRs (Eichinger *et al.* 2005; Sucgang *et al.* 2011; Heidel *et al.* 2011). Classically the GPCRs are
819 categorized into six classes based on sequence and function (Kolakowski 1994; Bockaert and Pin

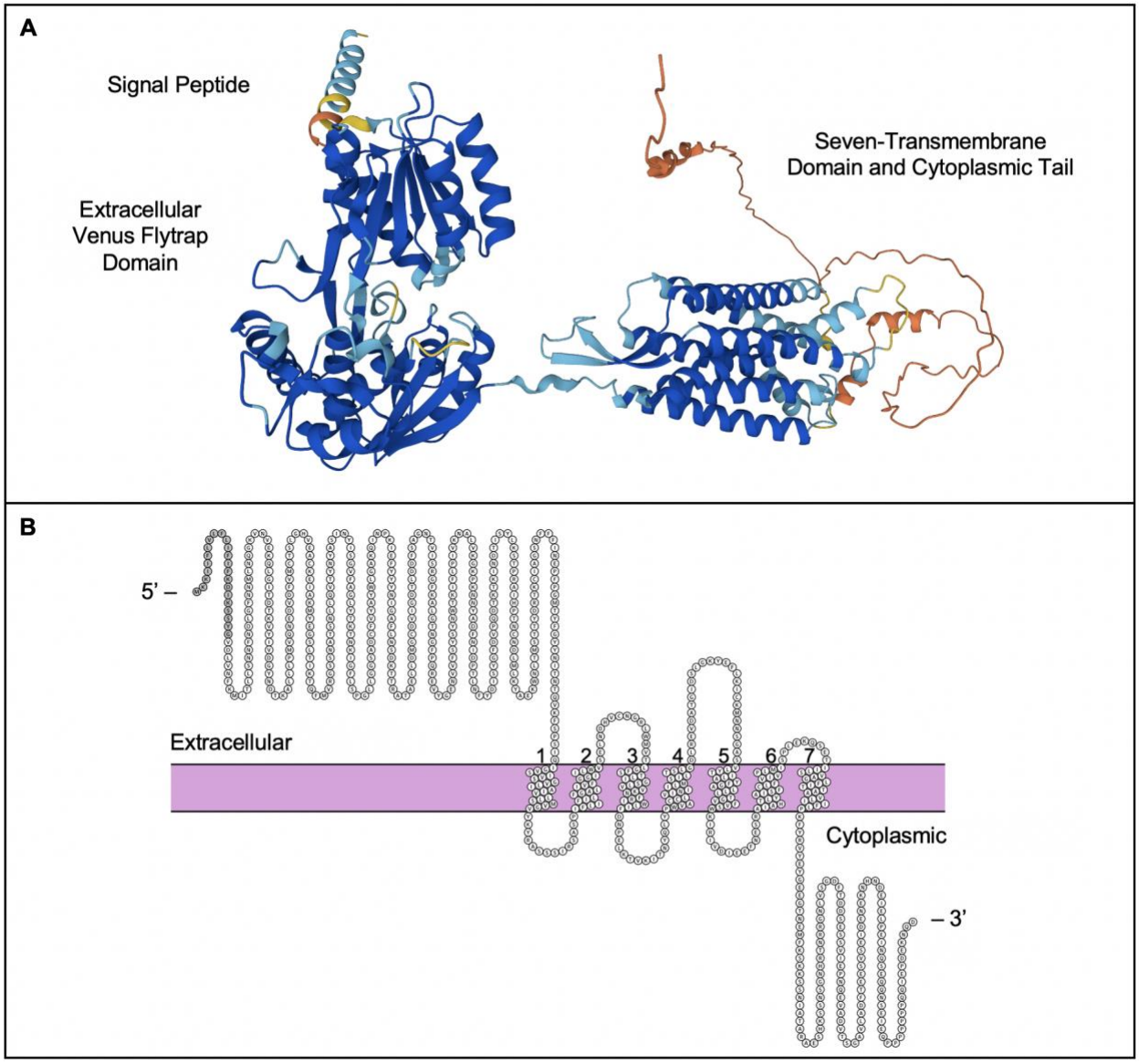
820 1999); *D. discoideum* has representative members of most classes (Prabhu and Eichinger 2006).
821 Similar to those in vertebrates, the GPCRs in *D. discoideum* are involved in a diversity of
822 biological processes, including significant roles in development. Most notably, members of the
823 class E GPCRs include the receptors of cAMP (cARs). The chemoattractant cAMP is pivotal in the
824 initiation and coordination of multicellular development in *D. discoideum* (Saran *et al.* 2002;
825 Loomis 2014). Perhaps it is not surprising then that the cAMP receptors were the first GPCRs to
826 be identified (Klein *et al.* 1988) in *D. discoideum* and remain the most extensively studied.
827
828 Some members of the class C GPCRs (which includes *grlG*) in *D. discoideum* have also been
829 implicated in development. The *D. discoideum* genome has 17 members of the class C GPCRs,
830 known as glutamate receptor-like (grl) proteins (Eichinger *et al.* 2005). For most of the group,
831 the ligands, effectors and even the specific signaling pathways remain unknown. A great deal of
832 progress has been made over the past few decades especially after the completion of the *D.*
833 *discoideum* genome in 2005 (Eichinger *et al.* 2005). At least four of the glutamate receptor-like
834 proteins have been described as having roles in development including *grlA* (Prabhu *et al.*
835 2007a; Anjard *et al.* 2009), *grlB* (Wu and Janetopoulos 2013), *grlE* (Anjard and Loomis 2006; Wu
836 and Janetopoulos 2013) and *grlJ* (Prabhu *et al.* 2007b). Moreover, Prabhu *et al.* (2007b) also
837 found that all 17 are expressed throughout development, some with increased expression early
838 in aggregation (*grlA, D, E, J, M, N, Q*) and others with increased expression from the tight
839 aggregate stage and on (including *grlG, B, C, F, H, K, L, O, P, R*). Much remains to be learned but
840 it is easy to conceive that *grlG,* like other members of this protein family in *D. discoideum*, is
841 also involved in cooperation during multicellular development.

842

843 We are aware of only one previous study that investigated the function of *grlG* but it was in the
844 context of predation rather than development. Upon exposure to the chemoattractant folic
845 acid (used by *D. discoideum* to chemotax toward and phagocytose bacteria) Pan *et al.* (2016)
846 identified increased phosphorylation at key serine residues in both *grlG* and *grlL.* Both *grlG* and
847 *grlL* were thus considered candidate folic acid receptors (they referred to them as folic acid
848 receptors ('far') *far2* and *far1,* respectively). However, only *far1* (*grlL*), and not *grlG,* was
849 required for eliciting the response to folic acid and the authors moved forward with the
850 characterization of only *far1* (*grlL*) (Pan *et al.* 2016, 2018). During the work that followed, they
851 noticed that the loss of *far1* (*grlL*) does not completely abolish the response to folic acid and
852 lipopolysaccharide (LPS) which suggests that other proteins are involved. That observation is in
853 line with some earlier studies which also suggested that there are two folic acid receptors, each
854 with a different binding affinity (de Wit and van Haastert 1985; Segall *et al.* 1988). But *grlG* is
855 not required for the folic acid response and there is only limited evidence to support any
856 involvement (Pan *et al.* 2016; Lamrabet *et al.* 2020). We have no reason to believe that the
857 parallel evolution of *grlG* that we have identified is related to the detection or phagocytosis of
858 bacterial prey. The *D. discoideum* AX4 lines were adapted to the laboratory growth and nutrient
859 conditions well before the start of experimental evolution by Kuzdzal-Fick *et al.* (2011). Instead,
860 the primary novelty of the experimental environment was the extreme low-relatedness that
861 favored cheaters.

862

863 The number of large-effect variants in *grlG* suggests that the loss of *grlG* was adaptive under
864 the experimental conditions of low relatedness. What is less clear from our results is why only
865 the variants in the 5' region of *grlG* are significantly associated with non-fruiting and those in
866 the 3' region are not. To investigate, we viewed the amino acid sequence topology and
867 predicted protein structure to investigate other potential explanations (Figure 5B). *GrlG* shares
868 the same major features found in most class C GPCRs including a long 5' extracellular domain, a
869 seven-transmembrane domain toward the 3' end followed by an intracellular C-terminal tail
870 (Figure 5B) (Pin *et al.* 2003; Rosenbaum *et al.* 2009). Ligand binding in class C GPCRs typically
871 occurs via a Venus flytrap domain (similar to the ancestral periplasmic binding proteins of
872 bacteria) situated in the 5' extracellular domain (Cao *et al.* 2009; Chun *et al.* 2012). Upon
873 binding its ligand, conformational changes throughout the seven-transmembrane domain lead
874 to the activation of G proteins (or other effectors) at the intracellular C-terminus to induce
875 signaling inside the cell (Rosenbaum *et al.* 2009). Based on sequence homology and the
876 predicted folding structure, *grlG* appears to function similarly. In the predicted structure
877 generated by AlphaFold (Jumper *et al.* 2021; Varadi *et al.* 2022), the extracellular region can be
878 seen folded (with high confidence) into a Venus flytrap structure with a clearly visible cleft, or
879 ligand binding pocket (Figure 5A).
880
881



882

883 **Figure 5. Predicted structure and sequence topology of *D. discoideum grlG.* A**. Predicted

884 folding structure of *grlG* generated by AlphaFold (Jumper *et al.* 2021; Varadi *et al.* 2022). Colors

885 indicate the model confidence in the local accuracy; dark blue is very high confidence, light blue

886 is confident, yellow is low confidence and orange is very low confidence. This structure can be

887 be viewed interactively with the 3D viewer on the AlphaFold Protein Structure Database

888 (https://alphafold.ebi.ac.uk/entry/Q75JP4). **B.** The sequence topology of *grlG* generated in

889 Protter (Omasits *et al.* 2014). Extracellular regions are shown above the membrane (in purple)

890 and the cytoplasmic regions below. The helices of the seven-transmembrane domain are
891 numbered one through seven. The 5' amino acids in grey indicate the signal peptide.
892
893
894 According to the amino acid sequence topology**,** each of the four variants that we screened for
895 in the 5' region would result in early protein truncation, either upstream of or within the
896 extracellular binding domain. It is not surprising that the loss of the binding domain, such as in
897 our 5' variants, would render a receptor non-functional. Similar results have been
898 experimentally demonstrated for the homologous protein, *grlL* (*far1*). Pan *et al.* (2018)
899 generated a *grlL* (*far1*) mutant missing only the extracellular binding domain (the remainder of
900 the protein was intact) to demonstrate that the mutant was unable to elicit a response to its
901 ligands. But while we might predict that the large impact variants in the transmembrane
902 domain, particularly those also resulting in early protein truncation, would lead to similar
903 outcomes, our data do not support a significant association between any of the 3' variants and
904 non-fruiting. Additional work is needed to understand how *D. discoideum* cells handle aberrant
905 *grlG* mRNAs and proteins. But even if we assume that aberrant and truncated proteins are not
906 degraded, given the coordinated functionality between extracellular ligand binding and the
907 resultant intracellular signaling, the loss of either end of this transmembrane protein receptor
908 would seem to disable signaling. Moreover, the variants in both regions of *grlG* were positively
909 selected and increased in abundance under the experimental conditions. Although it is difficult
910 to speculate given so many unknowns, perhaps all variants in *grlG* (both 5' and 3') provided the

911 selective advantage of cheating, but only those that inhibit ligand binding are also likely to
912 result in non-fruiting when in isolation.
913
914 Another challenge to interpreting our results is that variants in *grlG* are apparently not
915 completely penetrant. While our results are broadly consistent with variants in 5' *grlG* causing
916 non-fruiting, we did identify one clone in line 21 that was still able to form a fruiting body
917 despite carrying a variant in 5' *grlG* that introduces a premature stop codon just before the
918 extracellular binding domain. Although human or technical error is a possibility, there also
919 many potential biological explanations. Some of the factors that can influence penetrance
920 include gene expression levels, modifier genes, the co-occurrence of one or more additional
921 variants, the environment or other non-genetic factors (Cooper *et al.* 2013). However, the lack
922 of penetrance we observed may not be explained until we know more about the gene,
923 particularly the ligand(s) it binds and the pathway(s) it is involved in. This brings our attention
924 to the complex and pleiotropic nature of GPCR signaling and highlights the challenge of
925 predicting variant impacts.
926
927 From the great deal of work with GPCRs in humans and disease in particular, it is clear that
928 mutations in membrane proteins can result in a huge variety of impacts. The effect of any
929 mutation will depend on the properties of the protein as well as the type and location of the
930 mutation (Schöneberg and Liebscher 2021; Zaucha *et al.* 2021). Some of the many potential
931 impacts of GPCR mutations include improper membrane trafficking or localization; alterations
932 to critical secondary or tertiary structures; inhibition or altered efficiency of binding ligand(s) or

933 interacting proteins and inhibition or altered efficiency of signaling through downstream
934 effectors. Further complicating matters, some GPCRs bind multiple ligands and/or have
935 multiple effectors. Likewise, one ligand can have multiple receptors. In *D. discoideum* for
936 example, the chemoattractant cAMP has four receptors (cARs 1–4) in the GPCR class E, each of
937 which has a different cAMP-binding affinity and expression pattern. The binding affinity of each
938 cAMP receptor is correlated with the stage of development during which it is expressed to
939 optimally meet the needs of the cell at that time (Kim *et al.* 1996).
940
941 Members of the class C glutamate receptor-like protein family (which includes *grlG*) in *D.*
942 *discoideum* also exemplify complexities of GPCR signaling. *GrlE*, which is the GABA receptor
943 during late development, is also bound by the competitive inhibitor of GABA, glutamate.
944 Depending on which ligand is bound, GABA or glutamate, *grlE* is coupled to a different G
945 protein subunit (Gα7 or Gα9, respectively) (Anjard and Loomis 2006). Sporulation is delayed
946 while glutamate is bound and then induced when displaced and bound by GABA, together
947 ensuring the appropriate timing of sporulation (Anjard and Loomis 2006; Anjard *et al.* 2009).
948 Another example of a glutamate receptor-like protein binding multiple ligands is the folic acid
949 receptor *far1* (otherwise known as *grlL*), which also recognizes lipopolysaccharide (LPS) to
950 mediate chemotaxis and phagocytosis of bacterial prey (Pan *et al.* 2016, 2018). And similar to
951 the previously described multiple receptors of cAMP, other ligands bind multiple glutamate
952 receptor-like proteins of class C as well. While *grlE* is the GABA receptor during late
953 development, *grlB* is instead the GABA receptor during early stages of development (Wu and
954 Janetopoulos 2013). There are also several examples in *D. discoideum* of potential redundancy

955 (or rather, the suspected involvement of multiple receptors) of glutamate receptor-like
956 proteins, wherein the loss of the focal receptor does not completely abolish the response as
957 would be expected if a single receptor were responsible for ligand recognition and response.
958 Examples in which the loss of the focal receptor does not completely abolish the response
959 include the autocrine proliferation repressor protein A (*aprA*) receptor, *grlH* (Tang *et al.* 2018);
960 the phenylthiourea receptor, *grlJ* (Robery *et al.* 2013); and the folic acid (and LPS) receptor, *far1*
961 (*grlL*) (Pan *et al.* 2016, 2018).
962
963 Regarding the potential redundancy of *far1* (*grlL*), as previously mentioned, the involvement of
964 *grlG* has been suggested (Pan *et al.* 2016; Lamrabet *et al.* 2020) but that has little relevance in
965 this study as we have no reason to believe that our results are predation related. However, we
966 do not rule out the possibility *grlG* may also interact with folic acid or another molecule with
967 overlapping roles in bacterial predation as well as in cooperation and fruiting body formation.
968 Given our understanding of the complex nature of GPCR signaling, even if *grlG* was found to be
969 involved in the response to bacterial prey, that would not contradict the evidence we present
970 here suggesting its role in cooperation. Instead, perhaps the knowledge that *grlG* and other
971 members of the glutamate receptor-like protein family sometimes show low levels of
972 overlapping functionality may help explain why the loss of *grlG* in our evolved clones did not
973 always lead to the total loss of cooperation and non-fruiting.
974
975
976 **Conclusion**

977 In this study we returned to cell lines previously evolved under low-relatedness by Kuzdzal-Fick
978 *et al.* (2011) to investigate the genomic changes underlying the observed evolution of non-
979 fruiting obligate cheaters. That study demonstrated how low relatedness in a natural
980 population would lead to a collapse of multicellularity and with it, the advantages of fruiting
981 body formation and spore dispersal. In this study we used whole-genome sequencing and
982 variant analysis of those experimentally evolved *D. discoideum* low-relatedness cell lines. We
983 identified parallel evolution of the orphan GPCR *grlG* as well as another 28 variants in other
984 genes. The genes that we identified with variants are spread throughout the genome and do
985 not share any significantly enriched functions or other annotations. In addition, none of these
986 genes have been previously implicated in cheating or cooperation. The lack of overlap among
987 these genes alone, nor with previously implicated cheater genes, is further testament to the
988 complexity of the social cycle and aggregative multicellularity in *D. discoideum* and further
989 highlights that these processes are orchestrated by a large diversity of genes and pathways.
990
991 The orphan GPCR *grlG* acquired variants of moderate to high impact in more than half of the
992 experimental lines suggesting that the loss of *grlG* was adaptive under the experimental
993 conditions of low-relatedness. By screening additional clones from the evolved lines, we
994 identified a significant correlation between the complete loss of cooperation (inability to form
995 the social stage, a fruiting body) and the presence of variants in the 5' region of *grlG* containing
996 the signal peptide and extracellular binding domain. Our results suggest that *grlG* is one of now
997 two genes (second to *fbxA* (Ennis *et al.* 2000)) to have been implicated in obligate non-fruiting
998 cheating in *D. discoideum*. Although future work is needed to identify the interacting molecules

999 and the precise role of *grlG*, this work adds *grlG* to the growing list of GPCRs in the *D.*
1000 *discoideum* genome to have a role in multicellular development. This is noteworthy because it
1001 brings us one step closer to deorphanization of the ever-important GPCRs and highlights the
1002 need to expand research efforts to characterize the remaining GPCRs in *D. discoideum.*
1003
1004
1005 ***Acknowledgments***
1006 We thank members of the Queller-Strassmann laboratory for helpful advice on this research
1007 and Jennie Kuzdzal-Fick for making her evolved lines available. This material is based upon work
1008 supported by the National Science Foundation under grant numbers IOS 16-56756 and DEB 17-
1009 53743. Additional funding came from a pilot grant from the McDonnell Genome Institute.
1010
1011
1012 ***Supplemental Files***
1013 **Supplemental Methods**
1014 **Supplemental Tables**
1015 **Additional File 1** Results of the limited screening of clones for variants in genes other than *grlG*
1016 **Supplemental Material 1** Example images of non-fruiting clone morphology
1017 **Supplemental Material 2** Variant calls that survived filtration from each of the three variant
1018 callers (GATK, Freebayes and Delly), in VCF format
1019 **Supplemental Material 3** *Dictyostelium discoideum* gene annotations downloaded from NCBI
1020 on 10-25-19

1021 **Supplemental Material 4** Tables showing the resulting number of variants after each step of

1022 filtration of the VCF files independently generated by each caller.

1023 **Supplemental Material 5** Results of the functional annotation clustering with a high

1024 classification efficiency performed for the 29 genes with SNPs using the online tool DAVID

1025 (https://david.ncifcrf.gov/)

1026

1027

1028 **References**

1029 Anjard, C., and W. F. Loomis, 2006 GABA induces terminal differentiation of *Dictyostelium*

1030 through a $GABA_B$ receptor. Development 133: 2253–2261.

1031 Anjard, C., Y. Su, and W. F. Loomis, 2009 Steroids initiate a signaling cascade that triggers rapid

1032 sporulation in *Dictyostelium*. Development 136: 803–812.

1033 Barrick, J. E., and R. E. Lenski, 2013 Genome dynamics during experimental evolution. Nat. Rev.

1034 Genet. 14: 827–839.

1035 Bateman, A., M. J. Martin, S. Orchard, M. Magrane, R. Agivetova *et al.*, 2021 UniProt: The

1036 universal protein knowledgebase in 2021. Nucleic Acids Res. 49: D480–D489.

1037 Benabentos, R., S. Hirose, R. Sucgang, T. Curk, M. Katoh *et al.*, 2009 Polymorphic members of

1038 the lag gene family mediate kin discrimination in *Dictyostelium*. Curr. Biol. 19: 567–572.

1039 Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: A practical and

1040 powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57: 289–300.

1041 Bockaert, J., and J. P. Pin, 1999 Molecular tinkering of G protein-coupled receptors: An

1042 evolutionary success. EMBO J. 18: 1723–1729.

1043 Bonner, J. T., 1998 The origins of multicellularity. Integr. Biol. Issues News Rev. 1: 27–36.

1044 Bozzaro, S., 2019 The past, present and future of *Dictyostelium* as a model system. Int. J. Dev.
1045 Biol. 63: 321–331.

1046 Buttery, N. J., C. N. Jack, B. Adu-Oppong, K. T. Snyder, C. R. L. Thompson *et al.*, 2012 Structured
1047 growth and genetic drift raise relatedness in the social amoeba *Dictyostelium*
1048 *discoideum*. Biol. Lett. 8: 794–797.

1049 Buttery, N. J., D. E. Rozen, J. B. Wolf, and C. R. L. Thompson, 2009 Quantification of social
1050 behavior in *D. discoideum* reveals complex fixed and facultative strategies. Curr. Biol. 19:
1051 1373–1377.

1052 Buttery, N. J., J. Smith, D. C. Queller, and J. E. Strassmann, 2013 Measuring cheating, fitness,
1053 and segregation in *Dictyostelium discoideum*, pp. 231–248 in *Dictyostelium discoideum*
1054 *Protocols*, Humana Press, Totowa, NJ.

1055 Cameron, D. L., L. Di Stefano, and A. T. Papenfuss, 2019 Comprehensive evaluation and
1056 characterisation of short read general-purpose structural variant calling software. Nat.
1057 Commun. 10: 1–11.

1058 Cao, J., S. Huang, J. Qian, J. Huang, L. Jin *et al.*, 2009 Evolution of the class C GPCR Venus flytrap
1059 modules involved positive selected functional divergence. BMC Evol. Biol. 9: 1–12.

1060 Cavender, J. C., and K. B. Raper, 1965 The Acrasieae in nature. I. Isolation. Am. J. Bot. 52: 294–
1061 296.

1062 Charette, S. J., and P. Cosson, 2004 Preparation of genomic DNA from *Dictyostelium discoideum*
1063 for PCR analysis. BioTechniques 36: 574–575.

1064 Chun, L., W. H. Zhang, and J. F. Liu, 2012 Structure and ligand recognition of class C GPCRs. Acta
1065 Pharmacol. Sin. 33: 312–323.
1066 Cooper, D. N., M. Krawczak, C. Polychronakos, C. Tyler-Smith, and H. Kehrer-Sawatzki, 2013
1067 Where genotype is not predictive of phenotype: Towards an understanding of the
1068 molecular basis of reduced penetrance in human inherited disease. Hum. Genet. 132:
1069 1077–1130.
1070 Cooper, G. A., and S. A. West, 2018 Division of labour and the evolution of extreme
1071 specialization. Nat. Ecol. Evol. 2: 1161–1167.
1072 Eichinger, L., J. A. Pachebat, G. Glöckner, M. A. Rajandream, R. Sucgang *et al.*, 2005 The genome
1073 of the social amoeba *Dictyostelium discoideum*. Nature 435: 43–57.
1074 Ellegren, H., 2004 Microsatellites: Simple sequences with complex evolution. Nat. Rev. Genet. 5:
1075 435–445.
1076 Ennis, H. L., D. N. Dao, S. U. Pukatzki, and R. H. Kessin, 2000 *Dictyostelium* amoebae lacking an
1077 F-box protein form spores rather than stalk in chimeras with wild type. Proc. Natl. Acad.
1078 Sci. 97: 3292–3297.
1079 Ennis, H. L., D. N. Dao, M. Y. Wu, and R. H. Kessin, 2003 Mutation of the *Dictyostelium fbxA*
1080 gene affects cell-fate decisions and spatial patterning. Protist 154: 419–429.
1081 Fey, P., P. Gaudet, T. Curk, B. Zupan, E. M. Just *et al.*, 2009 dictyBase—a *Dictyostelium*
1082 bioinformatics resource update. Nucleic Acids Res. 37: D515–D519.
1083 Flowers, J. M., S. I. Li, A. Stathos, G. Saxer, E. A. Ostrowski *et al.*, 2010 Variation, sex, and social
1084 cooperation: Molecular population genetics of the social amoeba *Dictyostelium*
1085 *discoideum*. PLoS Genet. 6: 1–14.

1086 Fortunato, A., D. C. Queller, and J. E. Strassmann, 2003a A linear dominance hierarchy among
1087 clones in chimeras of the social amoeba *Dictyostelium discoideum*. J. Evol. Biol. 16: 438–
1088 445.
1089 Fortunato, A., J. E. Strassmann, L. Santorelli, and D. C. Queller, 2003b Co-occurrence in nature
1090 of different clones of the social amoeba, *Dictyostelium discoideum*. Mol. Ecol. 12: 1031–
1091 1038.
1092 Foster, K. R., G. Shaulsky, J. E. Strassmann, D. C. Queller, and C. R. L. Thompson, 2004 Pleiotropy
1093 as a mechanism to stabilize cooperation. Nature 431: 693–696.
1094 Garrison, E., Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, and P. Prins, 2021 Vcflib and tools
1095 for processing the VCF variant call format. bioRxiv 2021.05.21.445151.
1096 Garrison, E., and G. Marth, 2012 Haplotype-based variant detection from short-read
1097 sequencing.
1098 Gilbert, O. M., K. R. Foster, N. J. Mehdiabadi, J. E. Strassmann, and D. C. Queller, 2007 High
1099 relatedness maintains multicellular cooperation in a social amoeba by controlling
1100 cheater mutants. Proc. Natl. Acad. Sci. U. S. A. 104: 8913–8917.
1101 Gilbert, O. M., J. E. Strassmann, and D. C. Queller, 2012 High relatedness in a social amoeba:
1102 The role of kin-discriminatory segregation. Proc. R. Soc. B Biol. Sci. 279: 2619–2624.
1103 Grosberg, R. K., and R. R. Strathmann, 2007 The evolution of multicellularity: A minor major
1104 transition? Annu. Rev. Ecol. Evol. Syst. 38: 621–654.
1105 Gruenheit, N., K. Parkinson, B. Stewart, J. A. Howie, J. B. Wolf *et al.*, 2017 A polychromatic
1106 “greenbeard” locus determines patterns of cooperation in a social amoeba. Nat.
1107 Commun. 8: 1–9.

1108 Hamilton, W. D., 1964 The genetical evolution of social behaviour. II. J. Theor. Biol. 7: 17–52.

1109 Hardin, G., 1968 The tragedy of the commons. Science 162: 1243–1248.

1110 Hauser, A. S., M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, and D. E. Gloriam, 2017 Trends
1111 in GPCR drug discovery: New agents, targets and indications. Nat. Rev. Drug Discov. 16:
1112 829–842.

1113 Heidel, A. J., H. M. Lawal, M. Felder, C. Schilde, N. R. Helps *et al.*, 2011 Phylogeny-wide analysis
1114 of social amoeba genomes highlights ancient origins for complex intercellular
1115 communication. Genome Res. 21: 1882–1891.

1116 Hirose, S., R. Benabentos, H. I. Ho, A. Kuspa, and G. Shaulsky, 2011 Self-recognition in social
1117 amoebae is mediated by allelic pairs of tiger genes. Science 333: 467–470.

1118 Hirose, S., B. Santhanam, M. Katoh-Kurosawa, G. Shaulsky, and A. Kuspa, 2015 Allorecognition,
1119 via *tgrB1* and *tgrC1*, mediates the transition from unicellularity to multicellularity in the
1120 social amoeba *Dictyostelium discoideum*. Development 142: 3561–3570.

1121 Ho, H. I., S. Hirose, A. Kuspa, and G. Shaulsky, 2013 Kin recognition protects cooperators against
1122 cheaters. Curr. Biol. 23: 1590–1595.

1123 Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009a Bioinformatics enrichment tools: Paths
1124 toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:
1125 1–13.

1126 Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009b Systematic and integrative analysis of
1127 large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44–57.

1128 Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov *et al.*, 2021 Highly accurate protein
1129 structure prediction with AlphaFold. Nature 596: 583–589.

1130 Kessin, R. H., 2001 *Dictyostelium: evolution, cell biology, and the development of*
1131 *multicellularity*. Cambridge University Press.
1132 Khare, A., and G. Shaulsky, 2010 Cheating by exploitation of developmental prestalk patterning
1133 in *Dictyostelium discoideum*. PLoS Genet. 6: e1000854.
1134 Kim, J., P. V. Haastert, and P. N. Devreotes, 1996 Social senses: G-protein-coupled pathways in
1135 *Dictyostelium*. Chem. Biol. 3: 239–243.
1136 Klein, P. S., T. J. Sun, C. L. Saxe, A. R. Kimmel, L. Ronald *et al.*, 1988 A chemoattractant receptor
1137 controls development in *Dictyostelium discoideum*. Science 241: 1467–1472.
1138 Knoll, A. H., 2011 The multiple origins of complex multicellularity. Annu. Rev. Earth Planet. Sci.
1139 39: 217–239.
1140 Kolakowski, L. F., 1994 GCRDb: A G-protein-coupled receptor database. Receptors Channels 2:
1141 1–7.
1142 Kosugi, S., Y. Momozawa, X. Liu, C. Terao, M. Kubo *et al.*, 2019 Comprehensive evaluation of
1143 structural variation detection algorithms for whole genome sequencing. Genome Biol.
1144 20: 8–11.
1145 Kuspa, A., and W. F. Loomis, 1992 Tagging developmental genes in *Dictyostelium* by restriction
1146 enzyme-mediated integration of plasmid DNA. Proc. Natl. Acad. Sci. U. S. A. 89: 8803–
1147 8807.
1148 Kuzdzal-Fick, J. J., S. A. Fox, J. E. Strassmann, and D. C. Queller, 2011 High relatedness is
1149 necessary and sufficient to maintain multicellularity in *Dictyostelium*. Science 334:
1150 1548–1551.

1151 Lagerström, M. C., and H. B. Schiöth, 2008 Structural diversity of G protein-coupled receptors
1152 and significance for drug discovery. Nat. Rev. Drug Discov. 7: 339–357.
1153 Lamrabet, O., A. Melotti, F. Burdet, N. Hanna, J. Perrin *et al.*, 2020 Transcriptional responses of
1154 *Dictyostelium discoideum* exposed to different classes of bacteria. Front. Microbiol. 11:
1155 1–12.
1156 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping
1157 and population genetical parameter estimation from sequencing data. Bioinformatics
1158 27: 2987–2993.
1159 Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1160 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
1161 transform. Bioinformatics 25: 1754–1760.
1162 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map
1163 format and SAMtools. Bioinformatics 25: 2078–2079.
1164 Loomis, W. F., 2014 Cell signaling during development of *Dictyostelium*. Dev. Biol. 391: 1–16.
1165 Loomis, W. F., 2015 Genetic control of morphogenesis in *Dictyostelium*. Dev. Biol. 402: 146–
1166 161.
1167 Madgwick, P. G., B. Stewart, L. J. Belcher, C. R. L. Thompson, and J. B. Wolf, 2018 Strategic
1168 investment explains patterns of cooperation and cheating in a microbe. Proc. Natl. Acad.
1169 Sci. U. S. A. 115: E4823–E4832.
1170 Mahmoud, M., N. Gobet, D. I. Cruz-Dávalos, N. Mounier, C. Dessimoz *et al.*, 2019 Structural
1171 variant calling: The long and the short of it. Genome Biol. 20: 1–14.

1172 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis
1173 toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
1174 Genome Res. 20: 1297–1303.
1175 McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al.*, 2016 The Ensembl Variant Effect
1176 Predictor. Genome Biol. 17: 122.
1177 Mehdiabadi, N. J., C. N. Jack, T. T. Farnham, T. G. Platt, S. E. Kalla *et al.*, 2006 Kin preference in a
1178 social microbe. Nature 442: 881–882.
1179 Mehdiabadi, N. J., M. R. Kronforst, D. C. Queller, and J. E. Strassmann, 2009 Phylogeny,
1180 reproductive isolation and kin recognition in the social amoeba *Dictyostelium*
1181 *purpureum*. Evol. Int. J. Org. Evol. 63: 542–8.
1182 Noh, S., L. Christopher, J. E. Strassmann, and D. C. Queller, 2020 Wild *Dictyostelium discoideum*
1183 social amoebae show plastic responses to the presence of nonrelatives during
1184 multicellular development. Ecol. Evol. 10: 1119–1134.
1185 Noh, S., K. S. Geist, X. Tian, J. E. Strassmann, and D. C. Queller, 2018 Genetic signatures of
1186 microbial altruism and cheating in social amoebas in the wild. Proc. Natl. Acad. Sci. U. S.
1187 A. 115: 3096–3101.
1188 de Oliveira, J. L., A. C. Morales, B. Stewart, N. Gruenheit, J. Engelmoer *et al.*, 2019 Conditional
1189 expression explains molecular evolution of social genes in a microbe. Nat. Commun. 10:
1190 1–12.
1191 Omasits, U., C. H. Ahrens, S. Müller, and B. Wollscheid, 2014 Protter: Interactive protein feature
1192 visualization and integration with experimental proteomic data. Bioinformatics 30: 884–
1193 886.

1194 Ostrowski, E. A., 2019 Enforcing cooperation in the social amoebae. Curr. Biol. 29: R474–R484.

1195 Pan, M., M. P. Neilson, A. M. Grunfeld, P. Cruz, X. Wen *et al.*, 2018 A G-protein-coupled
1196 chemoattractant receptor recognizes lipopolysaccharide for bacterial phagocytosis. PLoS
1197 Biol. 16: 1–20.

1198 Pan, M., X. Xu, Y. Chen, and T. Jin, 2016 Identification of a chemoattractant G-protein-coupled
1199 receptor for folic acid that controls both chemotaxis and phagocytosis. Dev. Cell 36:
1200 428–439.

1201 Pin, J. P., T. Galvez, and L. Prézeau, 2003 Evolution, structure, and activation mechanism of
1202 family 3/C G-protein-coupled receptors. Pharmacol. Ther. 98: 325–354.

1203 Prabhu, Y., and L. Eichinger, 2006 The *Dictyostelium* repertoire of seven transmembrane
1204 domain receptors. Eur. J. Cell Biol. 85: 937–946.

1205 Prabhu, Y., S. Mondal, L. Eichinger, and A. A. Noegel, 2007a A GPCR involved in post aggregation
1206 events in *Dictyostelium discoideum*. Dev. Biol. 312: 29–43.

1207 Prabhu, Y., R. Müller, C. Anjard, and A. A. Noegel, 2007b *GrlJ*, a *Dictyostelium* $GABA_B$-like
1208 receptor with roles in post-aggregation development. BMC Dev. Biol. 7: 1–14.

1209 Queller, D. C., 2000 Relatedness and the fraternal major transitions. Philos. Trans. R. Soc. B Biol.
1210 Sci. 355: 1647–1655.

1211 Queller, D. C., E. Ponte, S. Bozzaro, and J. E. Strassmann, 2003 Single-gene greenbeard effects in
1212 the social amoeba *Dictyostelium discoideum*. Science 299: 105–106.

1213 Queller, D. C., and J. E. Strassmann, 2018 Evolutionary conflict. Annu. Rev. Ecol. Evol. Syst. 49:
1214 73–93.

1215 Rausch, T., T. Zichner, A. Schlattl, A. M. Stütz, V. Benes *et al.*, 2012 DELLY: Structural variant
1216 discovery by integrated paired-end and split-read analysis. Bioinformatics 28: i333–i339.
1217 Robery, S., R. Tyson, C. Dinh, A. Kuspa, A. A. Noegel *et al.*, 2013 A novel human receptor
1218 involved in bitter tastant detection identified using *Dictyostelium discoideum*. J. Cell Sci.
1219 126: 5465–5476.
1220 Robinson, J. T., H. Thorvaldsdóttir, A. M. Wenger, A. Zehir, and J. P. Mesirov, 2017 Variant
1221 review with the integrative genomics viewer. Cancer Res. 77: e31–e34.
1222 Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011
1223 Integrative genomics viewer. Nat. Biotechnol. 29: 24–26.
1224 Rosenbaum, D. M., S. G. F. Rasmussen, and B. K. Kobilka, 2009 The structure and function of G-
1225 protein-coupled receptors. Nature 459: 356–363.
1226 Santorelli, L. A., A. Kuspa, G. Shaulsky, D. C. Queller, and J. E. Strassmann, 2013 A new social
1227 gene in *Dictyostelium discoideum*, *chtB*. BMC Evol. Biol. 13: 1–10.
1228 Santorelli, L. A., C. R. L. Thompson, E. Villegas, J. Svetz, C. Dinh *et al.*, 2008 Facultative cheater
1229 mutants reveal the genetic complexity of cooperation in social amoebae. Nature 451:
1230 1107–1110.
1231 Saran, S., M. E. Meima, E. Alvarez-Curto, K. E. Weening, D. E. Rozen *et al.*, 2002 cAMP signaling
1232 in *Dictyostelium*: Complexity of cAMP synthesis, degradation and detection. J. Muscle
1233 Res. Cell Motil. 23: 793–802.
1234 Sathe, S., N. Khetan, and V. Nanjundiah, 2014 Interspecies and intraspecies interactions in social
1235 amoebae. J. Evol. Biol. 27: 349–362.

1236 Schöneberg, T., and I. Liebscher, 2021 Mutations in G protein–coupled receptors: Mechanisms,
1237 pathophysiology and potential therapeutic approaches. Pharmacol. Rev. 73: 89–119.
1238 Scott, T. J., and D. C. Queller, 2019 Long-term evolutionary conflict, Sisyphean arms races, and
1239 power in Fisher's geometric model. Ecol. Evol. 9: 11243–11253.
1240 Sebé-Pedrós, A., B. M. Degnan, and I. Ruiz-Trillo, 2017 The origin of Metazoa: A unicellular
1241 perspective. Nat. Rev. Genet. 18: 498–512.
1242 Segall, J. E., A. A. Bominaar, E. Wallraff, and R. J. De Wit, 1988 Analysis of a *Dictyostelium*
1243 chemotaxis mutant with altered chemoattractant binding. J. Cell Sci. 91 (Pt 4): 479–489.
1244 smith, J., D. C. Queller, and J. E. Strassmann, 2014 Fruiting bodies of the social amoeba
1245 *Dictyostelium discoideum* increase spore transport by *Drosophila*. BMC Evol. Biol. 14: 1–
1246 5.
1247 smith, J., J. E. Strassmann, and D. C. Queller, 2016 Fine-scale spatial ecology drives kin selection
1248 relatedness among cooperating amoebae. Evolution 70: 848–859.
1249 Srivastava, S., A. K. Avvaru, D. T. Sowpati, and R. K. Mishra, 2019 Patterns of microsatellite
1250 distribution across eukaryotic genomes. BMC Genomics 20: 1–14.
1251 Strassmann, J. E., O. M. Gilbert, and D. C. Queller, 2011 Kin discrimination and cooperation in
1252 microbes. Annu. Rev. Microbiol. 65: 349–67.
1253 Strassmann, J. E., and D. C. Queller, 2011 Evolution of cooperation and control of cheating in a
1254 social microbe. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl: 10855–62.
1255 Strassmann, J. E., Y. Zhu, and D. C. Queller, 2000 Altruism and social cheating in the social
1256 amoeba *Dictyostelium discoideum*. Nature 408: 965–967.

1257 Sucgang, R., A. Kuo, X. Tian, W. Salerno, A. Parikh *et al.*, 2011 Comparative genomics of the
1258 social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. Genome Biol.
1259 12: R20.
1260 Tan, A., G. R. Abecasis, and H. M. Kang, 2015 Unified representation of genetic variants.
1261 Bioinformatics 31: 2202–2204.
1262 Tang, Y., Y. Wu, S. E. Herlihy, F. J. Brito-Aleman, J. H. Ting *et al.*, 2018 An autocrine proliferation
1263 repressor regulates *Dictyostelium discoideum* proliferation and chemorepulsion using
1264 the G protein-coupled receptor *grlH*. mBio 9: e02443-17.
1265 Tarasov, A., A. J. Vilella, E. Cuppen, I. J. Nijman, and P. Prins, 2015 Sambamba: Fast processing
1266 of NGS alignment formats. Bioinformatics 31: 2032–2034.
1267 Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett *et al.*, 2012 The
1268 molecular diversity of adaptive convergence. Science 335: 457–461.
1269 Thompson, C. R. L., Q. Fu, C. Buhay, R. R. Kay, and S. Gad, 2004 A bZIP/bRLZ transcription factor
1270 required for DIF signaling in *Dictyostelium*. Development 131: 513–523.
1271 Tian, X., J. E. Strassmann, and D. C. Queller, 2011 Genome nucleotide composition shapes
1272 variation in simple sequence repeats. Mol. Biol. Evol. 28: 899–909.
1273 Tóth, G., Z. Gáspári, and J. Jurka, 2000 Microsatellites in different eukaryotic genomes: Surveys
1274 and analysis. Genome Res. 10: 967–981.
1275 Tsutsui, N. D., 2004 Scents of self: The expression component of self/non-self recognition
1276 systems. Ann. Zool. Fenn. 41: 713–727.

1277 Van den Bergh, B., T. Swings, M. Fauvart, and J. Michiels, 2018 Experimental design, population
1278 dynamics, and diversity in microbial experimental evolution. Microbiol. Mol. Biol. Rev.
1279 82: e00008-18.
1280 Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia *et al.*, 2022 AlphaFold Protein
1281 Structure Database: Massively expanding the structural coverage of protein-sequence
1282 space with high-accuracy models. Nucleic Acids Res. 50: D439–D444.
1283 West, S. A., A. S. Griffin, A. Gardner, and S. P. Diggle, 2006 Social evolution theory for
1284 microorganisms. Nat. Rev. Microbiol. 4: 597–607.
1285 de Wit, R. J. W., and P. J. M. van Haastert, 1985 Binding of folates to *Dictyostelium discoideum*
1286 cells. Demonstration of five classes of binding sites and their interconversion. Biochim.
1287 Biophys. Acta Biomembr. 814: 199–213.
1288 Wolf, J. B., J. A. Howie, K. Parkinson, N. Gruenheit, D. Melo *et al.*, 2015 Fitness trade-offs result
1289 in the illusion of social success. Curr. Biol. 25: 1086–1090.
1290 Wu, Y., and C. Janetopoulos, 2013 Systematic analysis of γ-aminobutyric acid (GABA)
1291 metabolism and function in the social amoeba *Dictyostelium discoideum*. J. Biol. Chem.
1292 288: 15280–15290.
1293 Zaucha, J., M. Heinzinger, A. Kulandaisamy, E. Kataka, Ó. L. Salvádor *et al.*, 2021 Mutations in
1294 transmembrane proteins: Diseases, evolutionary insights, prediction and comparison
1295 with globular proteins. Brief. Bioinform. 22:.
1296
1297
1298

1299 **Table 1. Number of SNPs and average read depth by sample.** Whole-genome sequencing and
1300 variant analysis resulted in 38 SNPs called in 25 of the 47 samples and distributed throughout
1301 17 of the 24 evolved lines. This table indicates the number of SNPs called for each sample. The
1302 average mapped read depth was calculated for the raw reads normalized by the length of all
1303 annotated genes. Samples from each evolved line, sequenced as populations, are simply
1304 numbered 1 - 24. Clonal samples isolated from each of the evolved lines are named according
1305 to the population number hyphenated with the non-fruiting clone ID (*e.g.,* "NF1"). The non-
1306 fruiting clone IDs were retained for record keeping purposes; they are not related to the
1307 number of clones sequenced for a line.

Sample	SNP Count	Average Mapped Read Depth
1	0	128
1-NF2	0	43
2	1	137
3	0	146
3-NF1	1	127
4	0	171
5	0	161
5-NF2	3	40
6	1	152
7	4	229
7-NF2	5	44
8	0	165
8-NF1	0	89
9	1	220

9-NF3	1	24
10	1	151
10-NF3	1	23
11	2	116
11-NF1	0	37
12	1	215
12-NF1	1	97
13	1	165
13-NF2	1	70
14	0	127
14-NF3	0	22
15	0	176
15-NF1	0	69
16	0	277
16-NF2	3	118
17	3	123
18	0	191
18-NF1	3	30
19	0	319
19-NF1	0	80
20	0	231
20-NF1	1	66
20-NF3	1	32
21	1	272
21-NF1	3	40
21-NF2	5	45

22	3	291
23	0	305
23-NF1	0	25
23-NF2	0	43
24	0	222
24-NF1	1	32
1	0	128

Table 2. List of called SNPs and the impacted locus. Whole-genome sequencing and variant analysis resulted in this list of 38 bi-allelic SNPs impacting 29 different genes. For each SNP this table provides the precise location in the genome and the Dictybase gene ID (followed by gene name when available), followed by the called samples. Samples from each evolved line, sequenced as populations, are simply numbered 1–24. Clonal samples isolated from each of the evolved lines are named according to the population number hyphenated with the non-fruiting clone ID (*e.g.,* "NF1"). The last two columns of the table are the variant consequence and the estimated impact rating made by the Ensembl Variant Effect Predictor (VEP) (VEP 95.3).

Chrom:Site	**Gene ID**	**Called Sample(s)**	**VEP Consequence**	**VEP Impact**
NC_007087.3:2556092	DDB_G0269332	22	missense_var.	Moderate
NC_007087.3:3010110	DDB_G0270828	21-NF1	missense_var.	Moderate
NC_007087.3:3624901	DDB_G0270964	18-NF1	missense_var.	Moderate
NC_007087.3:3928085	DDB_G0269956(RTE)	21-NF2	upstream_gene_var.	Modifier
NC_007087.3:4563335	DDB_G0270834(*sgmA*)	5-NF2	missense_var.	Moderate

NC_007088.5:1741979	DDB_G0272244(*grlG*)	6	stop_gained	High
NC_007088.5:1741982	DDB_G0272244(*grlG*)	7 & 7-NF2	missense_var.	Moderate
NC_007088.5:1742043	DDB_G0272244(*grlG*)	11	missense_var.	Moderate
NC_007088.5:1742355	DDB_G0272244(*grlG*)	17	missense_var.	Moderate
NC_007088.5:1742362	DDB_G0272244(*grlG*)	13 & 13-NF2	missense_var.	Moderate
NC_007088.5:1742375	DDB_G0272244(*grlG*)	10 & 10-NF3	stop_gained	High
NC_007088.5:1742506	DDB_G0272244(*grlG*)	18-NF1	missense_var.	Moderate
NC_007088.5:1743478	DDB_G0272244(*grlG*)	2	missense_var.	Moderate
NC_007088.5:1743823	DDB_G0272244(*grlG*)	21 & 21-NF2	stop_gained	High
NC_007088.5:1743844	DDB_G0272244(*grlG*)	22	stop_gained	High
NC_007088.5:1761939	DDB_G0272484	16-NF2	missense_var.	Moderate
NC_007088.5:4313995	DDB_G0274875(*rnf160*)	16-NF2	stop_gained	High
NC_007088.5:6565220	DDB_G0276291	17	missense_var.	Moderate
NC_007088.5:6762467	DDB_G0276367	17	stop_gained	High
NC_007088.5:6787007	DDB_G0276529	7-NF2 & 16-NF2	missense_var.	Moderate
NC_007088.5:6871012	DDB_G0276553	7 & 7-NF2	upstream_gene_var.	Modifier
NC_007088.5:7954466	DDB_G0277481	21-NF2	missense_var.	Moderate
NC_007089.4:1029309	DDB_G0278531	24-NF1	stop_gained	High
NC_007089.4:1073554	DDB_G0278559	7 & 7-NF2	missense_var.	Moderate
NC_007089.4:1101637	DDB_G0278575	7 & 7-NF2	upstream_gene_var.	Modifier
NC_007089.4:3459101	DDB_G0280505(*tmem144A*)	5-NF2	missense_var.	Moderate
NC_007089.4:5126190	DDB_G0281923(*mrhA*)	21-NF1	downstream_gene_var.	Modifier
NC_007089.4:5671190	DDB_G0282355	12 & 12-NF1	upstream_gene_var.	Modifier
NC_007090.3:2549009	DDB_G0284845(*gxcC*)	3-NF1	upstream_gene_var.	Modifier
NC_007091.3:1993104	DDB_G0288805	22	stop_gained	High

NC_007091.3:2942727	DDB_G0289555(*arkA*)	21-NF1	missense_var.	Moderate
NC_007091.3:2996049	DDB_G0289583	11	missense_var.	Moderate
NC_007091.3:4205385	DDB_G0290523	21-NF2	synonymous_var.	Low
NC_007091.3:4793752	DDB_G0290943(*pks39*)	20-NF1 & 20-NF3	missense_var.	Moderate
NC_007091.3:4956884	DDB_G0291085(*gxcE*)	21-NF2	missense_var.	Moderate
NC_007091.3:905736	DDB_G0287967	18-NF1	downstream_gene_var.	Modifier
NC_007092.3:1299121	DDB_G0292198(RTE)	5-NF2	missense_var.	Moderate
NC_007092.3:1984519	DDB_G0292696(*colA*)	9 & 9-NF3	missense_var.	Moderate

1318