Evolutionary Adaptations of IRG1 Refines Itaconate Synthesis and Mitigates Innate Immunometabolism Trade-offs

Richard V. Szeligowski[1,2], Francois Miros [1,2], Andres Saez[1,2], Marisa DeCiucis[1,2], Gunter P. Wagner[2,3], and Hongying Shen[1,2,*]

[1]Department of Cellular and Molecular Physiology, Yale University School of Medicine, CT, USA

[2]Systems Biology Institute, Yale University, CT, USA

[3]Department of Evolutionary Biology, University of Vienna, Vienna, Austria

[*] **Correspondence: hongying.shen@yale.edu (H.S.)**

Abstract

Itaconate is an innate immune metabolite specifically produced in activated immune cells via the decarboxylation of *cis*-aconitate, an intermediate of the TCA cycle. By inhibiting succinate-related metabolism, itaconate exerts antimicrobial properties at the expense of potentially disrupting the hosts' own central energy metabolism, a double-edged dilemma of immunometabolism. To explore the evolutionary logic of itaconate biosynthesis, we investigated the evolutionary trajectory of IRG1, which encodes for *cis*-aconitate decarboxylase (CAD), the enzyme responsible for itaconate production. Phylogenetic analysis reveals a putative independent acquisition of metazoan and fungal IRG1 from prokaryotic sources. In metazoans, IRG1 underwent gene duplication and subsequently lost the mitochondrial targeting sequence (MTS), relocating CAD outside the mitochondrial matrix and therefore preventing direct inhibition of energy metabolism. In basal metazoans that contain IRG1, oysters and amphioxus, primitive IRG1 expression is also induced by innate immune stimuli, suggesting an already specialized role of itaconate for innate immune defense in early bilaterians. Our integrated *in silico* and experimental analysis highlight the molecular adaptations in IRG1, including subcellular relocation, that optimize itaconate production for innate immunity in resolving a fundamental trade-off in immunometabolism.

47 **Introduction**

48 Recent multispecies analysis reveals that enzyme biochemistry dedicated to modern eukaryotic
49 innate immunity has deep evolutionary roots, with a specialized role in the innate immunity of
50 ancient life. Examples include the cGAS-STING pathway in animals that shares ancestry with
51 the prokaryotic cyclic-oligonucleotide in anti-phage defense[1-4], and the eukaryotic peptidoglycan
52 amidase that digests the bacterial cell wall and was horizontally acquired from prokaryotic
53 antibacterial toxins[5]. We rationalize that tracing the evolutionary trajectory of the specialized
54 biochemistry in innate immunity might provide insights into the regulatory mechanisms that
55 shape and optimize the modern innate immune response compatible with the organismal fitness.
56

57 Itaconate, an innate immunometabolite with both antimicrobial and anti-inflammatory roles[6-8] is
58 an exemplar of specialized biochemistry in mammalian innate immunity. Itaconate is produced
59 via a single decarboxylation step by the enzyme *cis*-aconitate decarboxylase, the gene product of
60 the immune response gene (IRG1, also named ACOD1)[9], shunting away the central energy
61 metabolic pathway TCA cycle intermediate *cis*-aconitate at a high flux rate (Fig. 1A). Itaconate
62 is highly reactive due to the terminal double bond, the alkene group on the C2 position (Fig. 1A).
63 This reactivity enables antimicrobial activity via inhibiting isocitrate lyase in the glyoxylate
64 shunt, a key energy metabolic pathway exclusively found in prokaryotes[9-11], and vitamin B12-
65 dependent methylmalonyl-CoA mutase-mediated odd chain carbon utilization[12,13]. However, in
66 the host, itaconate also inhibits key mitochondrial matrix-localized energy metabolism including
67 succinyl-CoA ligase[14], succinate dehydrogenase[15,16] and B12-dependent methylmalonyl-CoA
68 mutase producing succinate[12], in addition to functioning as a Michael acceptor reacting to
69 nucleophilic sulfur group of the cysteine residue of KEAP1 activating NRF2[17] and that of the

70 antioxidant glutathione[18]. Itaconate can be completely degraded by a citramalyl-CoA lyase
71 activity (Fig. 1A) that was discovered for human CLYBL gene protecting the vitamin B12
72 cofactor[12]. The same activity of bacterial citE provided fitness (virulence) to the bacterial
73 pathogen by degrading itaconate to counter the host defense[19], which illustrates an evolutionary
74 arms race.
75
76 Taken together, itaconate's dual inhibition of the host and pathogen's central energy metabolic
77 pathway presents a "double-edged" dilemma in immunometabolism: the host benefits from the
78 elimination of pathogens, but also suffers an apparent detrimental effect via the disruption of the
79 organism's own energy metabolism.
80
81 One mechanism to alleviate this dilemma involves a spatiotemporal restricted transcriptional
82 regulation of IRG1, where its expression is exclusively restricted to macrophages in humans,
83 mice, and zebrafish[20]. Within macrophages, IRG1 transcription is strictly kept silent and only
84 upregulated upon innate immune stimulation by bacterial pathogens or LPS. Only a few studies
85 reported the induction of IRG1 in other cell types by other stimuli, including IRG1 paralog in
86 zebrafish that is induced in terminally differentiated keratinocytes upon bacterial infection[21] and
87 the induction of human IRG1 in terminally differentiated neurons upon zika virus infection[22].
88 This highly selective transcriptional hierarchy for IRG1 is evidently successful in resolving the
89 "double-edged" dilemma in modern vertebrates. However, other mechanisms to resolve this
90 dilemma may have arisen in distant species where transcriptional regulation is less sophisticated.
91

92 To explore the evolutionary logic of itaconate production, we investigated the taxonomic
93 distribution and the evolutionary trajectory of IRG1 *in silico*. We discovered that fungal IRG1
94 and metazoan IRG1 were likely independently acquired from prokaryotes. Whole genome
95 duplication early in vertebrate history, specifically in the gnathstome stem lineage, produced two
96 paralogous IRG1 genes and subsequent loss events in therian mammal lineages (marsupials and
97 placental mammals) enabled a precise regulation of itaconate biosynthesis by removing the
98 mitochondrial targeting sequence (MTS) in IRG1 that prevents a direct inhibition of
99 mitochondrial matrix bioenergetics. Our cell- and tissue-based studies in the most distantly
100 diverged metazoan species that contain IRG1 (paleo-IRG1), mollusks (eastern oyster,
101 *Crassostrea virginica*) and basal chordates (amphioxus, *Branchiostoma floridae*), identified a
102 transcriptional induction of IRG1 by innate immune inducers, suggesting an ancestrally
103 dedicated innate immune role for IRG1. Together, our molecular phylogeny and non-model
104 organism studies reveal a multilayered evolutionary strategy that tightens the regulations of
105 itaconate production. We speculate that similar evolutionary regulatory mechanisms may be
106 broadly utilized to shape toxic innate immune biochemistry and to resolve the "double-edged"
107 dilemma in immune defense.
108
109 **Results**
110 To understand the occurrence of itaconate production in nature, our initial literature survey
111 identified two enzymes that synthesize itaconate: the cis-aconitate decarboxylase (CAD) enzyme
112 in humans, mice[9], zebrafish[20], and the fungus *Aspergillus terreus*[23,24], and *trans*-aconitate
113 decarboxylase (TAD) in *Ustilago maydis*, a plant pathogenic fungus[25] (Fig. 1A). Primary
114 sequence analysis and domain searches suggest that all CAD sequences are comprised of the

115 MmgE/PrpD domain fold[26], a protein domain family that also contains the prokaryotic
116 methylcitrate dehydratase in the methylcitrate cycle[27]. The TAD enzymes are comprised of a
117 completely different protein fold that is shared by the adenylosuccinate lyase enzyme ADSL in
118 mammalian purine nucleotide biosynthesis. However, TAD enzymes appear to be exclusively
119 restricted to the fungi kingdom. For CAD, many fungal studies focus on the bioengineering role
120 of itaconate production for polymer biosynthesis[28], and the physiological role of *A. terreus* CAD
121 and *U. maydis* TAD is poorly understood ─ a potential function in immune defense may hint at
122 convergent evolution to produce itaconate starting from different protein folds[29].
123
124 Because CAD enzymes exhibit a broad taxonomic distribution spanning fungi and mammals, we
125 then focused on the evolutionary trajectory of CAD across the tree of life. Combined BLAST
126 searches (BLASTP and TBLASTN searches using human IRG1 protein sequence as a query) and
127 a MmgE/PrpD Pfam domain search returned protein sequences spanning all three domains of life
128 in bacteria, archaea and eukaryotes. A closer inspection of the eukaryotic domain suggests a
129 sporadic presence of IRG1 homologs in fungi, deuterostomes, and lophotrochozoans, but CAD is
130 notably absent in their sister clades including ecdysozoans (arthropods, nematodes and their
131 relatives) and plants (Fig. 1B). Six putative Drosophila sequences were determined to be
132 misannotated and thus excluded for further analysis (see methods). The MmgE/PrpD domain-
133 containing proteins are present in Basidiomycota and Ascomycota, the fungi taxa including *A.*
134 *terreus*, but are absent in other basal fungi taxa including Chytridiomycota and Zygomycota. The
135 patchiness of IRG1 occurrence in nature suggests that itaconate biosynthesis is not critical, and
136 thus dispensable for eukaryotic central energy metabolism and organismal fitness.
137

138 Molecular phylogenetic analysis of all prokaryotic and eukaryotic IRG1 homologs
139 (Supplementary File 1) highlighted a clear separation of the fungi and animal IRG1 sequences,
140 suggesting two independent acquisition events. We first confirmed that all the IRG1 sequences
141 contain the MmgE/PrpD domain and then manually added eight experimentally validated 2-
142 methylcitrate dehydratases including bacterial sequences from *E. coli*[30], *S. typhimurium*[31], *C.*
143 *necator*[32], *C. glutamicum*[33], *B. subtilis*[34], and two different homologs in *M. tuberculosis*[35-37] as
144 well as a fungal sequence from *Y. lipolytica*[38]. The most striking observation from the
145 phylogenetic analysis is that, despite similar enzymatic activities, the validated fungi *A. terreus*
146 CAD and the metazoan IRG1 sequences reside on two distant branches intercalated by
147 prokaryotic sequences with high bootstrapping values (>75) (Fig. 1C), strongly suggesting that
148 metazoan CAD/ IRG1 and the fungi CAD are independently acquired.
149
150 The absence of the CAD enzyme sequences in many fungal taxa and the close relationship of *A.*
151 *terreus* CAD sequence with the prokaryotic sequences in adjacent branches may indicate a
152 putative direct horizontal gene transfer (HGT) into the *Aspergillus* ancestor. The eight
153 experimentally validated 2-methylcitrate dehydratases are clustered on a separate branch from
154 both *A. terreus* CAD and metazoan IRG1, suggesting they are indeed different enzymes that can
155 be distinguished by sequence homology. Notably, the nearest clade to metazoan IRG1 orthologs
156 contain alphaproteobacterial sequences including one from the Rickettsiales genera, the bacterial
157 taxa considered to be the closest extant organism to eukaryotic mitochondria[39]. However, the
158 origin of prokaryotic clades where metazoans acquired IRG1 remains unknown.
159

160 The evolutionary trajectory of metazoan IRG1 is consistent with canonical metazoan phylogeny
161 and is composed of three distinct subbranches: a homolog in mollusks and basal chordates we
162 called paleo-IRG1 (pIRG1, Fig. 2, in gray) and two paralogous vertebrate groups that we
163 designated IRG1L (IRG1-like, Fig. 2, in blue) and IRG1 (Fig. 2, in red). The pIRG1 sequences
164 consist of invertebrate sequences exclusively in mollusks, the basal chordate amphioxus
165 (*Branchiostoma* sp.), and agnathans (lamprey, *Petromyzon marinus*) (Fig. 1C). A clear
166 divergence of the IRG1L and IRG1 branches in several gnathostome classes (in Actinopterygii,
167 Sarcopterygii, Amphibia, Aves, and Reptilia) and the conspicuous absence of IRG1L in therian
168 mammals prompted us to further inspect the taxonomic distribution and phylogeny of IRG1s in
169 metazoans.
170
171 We then focused on the metazoan tree and color-coded the sequences in the three subbranches
172 corresponding to pIRG1, IRG1 and IRG1L, by taxonomy (Fig. S1A). After analyzing the
173 taxonomic distribution of the IRG1 homologs, we confirmed that all IRG1 and IRG1L sequences
174 belong to jawed vertebrates and that each paralogous branch follows a canonical vertebrate
175 phylogeny. The presence of both IRG1 and IRG1L in the majority of jawed vertebrate taxa,
176 including bony fish, amphibians, reptiles, and birds and the exclusive presence of pIRG1 in
177 mollusks and agnathans strongly support that IRG1 and IRG1L are derived from the pIRG1
178 through the putative genome duplication event in the gnathostome stem-lineage[40].
179
180 Several taxon-specific local gene duplication events were observed, including two syntenic
181 pIRG1s in oysters (> 99.48% nucleotide sequence identity), two syntenic pIRG1s in lamprey (>

182 98.72% nucleotide sequence identity), four syntenic IRG1Ls in *Xenopus laevis* (Fig. 2 and Fig.
183 S1), and two IRG1Ls in salmonid fish likely due to known genome duplication in these taxa[41].
184
185 Cartilaginous fish genomes were found to only contain IRG1L, but lack IRG1 (Fig. S1A), a
186 result we further confirmed by a manual BLAST search in species including the skate
187 (*Amblyraja radiata*) and sharks (*Chilscyllium plagiosum, Rhincodon typus, Chiloscyllium*
188 *punctatum, Carcharodon carcharias, Scyliorhinus torazame, Callorhinchus milii and*
189 *Scyliorhinus canicula*).
190
191 The most striking observation in the taxa-annotated metazoan tree is a complete loss of IRG1L in
192 the metatherian and eutherian mammals that otherwise exhibit an expansion of IRG1 lineage
193 (Fig. S1A, pink), which contrasts with the presence of both IRG1 and IRG1L in the prototherian
194 platypus (*O. anatinus*; Figure S1A, arrows). To further confirm the difference, we manually
195 searched for IRG1 and IRG1L in major mammalian orders[42] by either BlastP or TblastN search
196 using human IRG1 and zebrafish IRG1L sequences (Fig. S1B). The absence of both IRG1 and
197 IRG1L in Hyracoidea and Xenarthra (Fig. S1B) is consistent with the dispensability of IRG1 for
198 organismal energy metabolism. Thus, all metatherians and eutherians contain only IRG1. Within
199 the mammalian clade, IRG1L is exclusive to prototheria, a clade of egg-laying mammals, and is
200 completely lost in metatheria and eutheria, the viviparous mammals (Fig. S1B), inviting further
201 analysis in the difference between IRG1L and IRG1.
202
203 Primary sequence inspection unexpectedly revealed a N-terminally extended amino acid
204 sequence that is frequently present in both IRG1L and pIRG1 sequences, but much less frequent

205 in the IRG1 sequences. Because the CAD enzyme directly acts on intermediates of the
206 mitochondrial matrix-localized TCA cycle, we then explored if these N-terminal extensions
207 might encode the canonical mitochondrial targeting sequence (MTS)[43], the peptide motif
208 directing proteins into the mitochondrial matrix. Indeed, investigating all metazoan sequences
209 using MTS prediction software TargetP revealed high-confidence MTS present in 34.56% pIRG1
210 and IRG1L sequences, and only 5 IRG1 sequences (<1.59%; Fig. S1 and Fig. 2).
211
212 To further establish the paralogous relationship between IRG1L and IRG1, we performed a
213 manual synteny analysis to probe the genomic context surrounding pIRG1, IRG1L and IRG1 in
214 the representative species from the major metazoan taxa that include well-studied model
215 organisms (Fig. 2). This subset of the tree retains both the branch topography (pIRG1, IRG1L
216 and IRG1) and the enrichment of MTS in the pIRG1 and IRG1L homologs (Fig. 2). We then
217 manually annotated and compared the coding genes in the genomic region harboring
218 pIRG1/IRG1L/IRG1. Although we did not detect obvious functionally related gene clustering[44],
219 the synteny surrounding IRG1 appears to have converged to a more highly conserved pattern
220 than that of IRG1L, both in the number of the syntenic genes and their directionality (Fig. 2).
221 Four closely examined genes syntenic to pIRG1 appear to contain the same protein domains as
222 the genes syntenic to IRG1L, but not the genes syntenic to IRG1, which we labeled based on
223 their conserved protein domains: VOC, YrdC, 2Fe (ii) oxy and GAT (Fig. 2). The exception is
224 the POZ domain protein that is found in the oyster synteny but are fused to BTB domain in the
225 human gene KCTD12 that is syntenic to human IRG1. Therefore, pIRG1 might have a more
226 similar synteny to IRG1L than to IRG1.
227

228 To experimentally study sub-mitochondrial localization, we applied protease K treatment of
229 mitochondrial fraction of HEK293 cells expressing zebrafish IRG1L that contains a predicted
230 MTS. Consistent with the *in silico* analysis, zebrafish IRG1L can only be digested in the
231 presence of a strong detergent Triton X-100, similar to a matrix protein marker HSP60,
232 confirming its matrix localization (Fig. 3A). In contrast, a recent super-resolution microscopy
233 and biochemistry using protease treatment identified that human IRG1 lacking MTS is not
234 localized to matrix, despite a mitochondrial enrichment[45]. We then engineered matrix-targeted
235 human IRG1 by fusing IRG1 with a model MTS, and demonstrated its matrix localization (Fig.
236 3B). Together, the matrix localization of zebrafish IRG1L might suggest a broader matrix
237 localization of pIRG1 and other IRG1Ls with MTS but not majority of IRG1 that lack MTS.
238 This loss of MTS in the IRG1 and loss of matrix-targeted IRG1L in the modern mammals might
239 confer a fitness advantage by redirecting IRG1 to cytoplasm and alleviating itaconate's direct
240 inhibition of the mitochondrial TCA cycle energy metabolism.
241
242 To confirm that pIRG1 and IRG1L are indeed cis-aconitate decarboxylase and are thus
243 implicated in innate immunity similar to human IRG1, we performed both *in vitro* enzymology
244 to study the recombinant zebrafish IRG1L protein, as well as *ex vivo* cellular and tissue-level
245 analysis of paleo-IRG1 in eastern oysters (*Crassostrea virginica*) and amphioxus
246 (*Branchiostoma floridae*).
247
248 First, *in vitro* enzymology using the zebrafish IRG1L recombinant protein confirmed the same
249 cis-aconitate decarboxylase activity that is reported for human IRG1 and *A. terreus* ACOD[26]. We
250 purified N-terminally His-tagged recombinant zebrafish IRG1L protein (aa 32- 466) lacking the

251 predicted mitochondrial targeting sequence (Fig. 4A). Incubating this recombinant protein with
252 aconitate *in vitro* produced itaconate in an enzyme- and time-dependent manner, detected by LC-
253 MS (Fig. 4B and 4C).
254
255 To probe a potentially ancient role of pIRG1 in innate immunity, we then analyzed pIRG1 in the
256 most distant metazoan known to contain the IRG1 gene, mollusks, specifically eastern oysters
257 (*C. virginica*). Under resting conditions, the pIRG1 mRNA can be detected from several major
258 oyster tissues including gills, mantle, stomach, muscles, and heart (Fig. 5A and 5B), however, a
259 strikingly high level of pIRG1 was detected in the hemocytes ($P < 0.0001$ relative to other
260 tissues; Fig. 5B), the cells in the oyster hemolymph extracted from the adductor muscles[46]. The
261 adult oysters contain a semi-open circulatory system in which the hemolymph (analogous to
262 blood) directly bathes other organ systems[47-49]. The oyster hemocytes exhibit dual functionality
263 in both nutrient acquisition and innate immune defense via phagocytic properties, resembling the
264 macrophages of the starfish larvae observed by Elie Metchnikoff[50,51].
265
266 To further explore pIRG1 in oyster innate immunity, we isolated and cultured pooled oyster
267 hemocytes and measured pIRG1 mRNA level in response to a panel of innate immune inducers:
268 heat-killed gram-negative bacterial pathogen *E. coli* (HKEB), heat-killed gram-positive bacterial
269 pathogen *S. aureus* (HKSA), zymosan (mimicking fungal infection), and Poly(I:C) (mimicking
270 viral infection) for 8 hours (Fig. S2). We found that oyster pIRG1 was minimally induced by
271 LPS (mimicking gram-negative bacterial infection), but was significantly upregulated by
272 poly(I:C) stimulation (Fig. S2 and Fig. 5E), and that Poly(I:C) induction upregulated pIRG1

273 expression up to 24-hr stimulation (Fig. 5D). The finding suggested an involvement of pIRG1 in
274 oyster innate immunity.
275
276 We then investigated pIRG1 in the basal chordate amphioxus *B. floridae*. Amphioxus, or
277 lancelets, belong to the subphylum cephalochordata, which is thought to have shared a common
278 ancestor with more derived chordates > 500 mya[52]. This 4-7 cm long, translucent, fish-like
279 animal contains a notochord, but not vertebrae, is considered a basal chordate and a model
280 organism to study origins of vertebrates[53]. Lacking circulatory cells, the cellular immunity in the
281 adult animals is provided by tissue-resident macrophage-like coelomocytes, as reported in the
282 intestine[54]. To probe amphioxus pIRG1 in innate immunity, the adult animals were stimulated
283 with LPS and Poly(I:C) by oral administration, and then recovered in salt water for 24 hrs prior
284 to tissue dissection and analysis for pIRG1 transcript (Fig. 6A). Paleo-IRG1 is barely detected in
285 the major tissues including the gills, hepatocecum, muscle and intestine in the control animals
286 (Fig. 6B), where the inconsistent detection with very large amplification cycle numbers (>35)
287 precluded an accurate normalization. In comparison, the pIRG1 transcript was dramatically
288 upregulated and consistently detected in all tissues after both LPS and Poly(I:C) stimulation (Fig.
289 6B). A transcriptional induction of amphioxus pIRG1 by LPS and Poly(I:C) supports the role of
290 pIRG1 and itaconate biochemistry in innate immunity early in metazoans.
291
292 **Discussion**
293 Our experimental studies of IRG1 in oysters and amphioxus, the most distant metazoan species
294 known to contain IRG1, revealed itaconate biosynthesis is enriched in the innate immune cells
295 (in oysters) and is induced upon innate immune stimulus (in oysters and amphioxus), suggesting

296 an ancient and yet already specialized role of itaconate in innate immunity. The induction of
297 oyster pIRG1 is consistent with the increased itaconate level reported in oysters upon marine
298 herpesvirus infection[55] and is also supported by the induced expression of pIRG1[56] and increased
299 itaconate[57,58] in another marine bivalve (mussels; *Perna canaliculus*) upon infection with a gram-
300 negative *Vibrio* strain. In these contexts, itaconate might play an antimicrobial function, as
301 itaconate inhibits a pathogenic marine *Vibrio* strain growth *in vitro* and disruption of central
302 carbon metabolism[59].
303
304 Taxonomic distribution and evolutionary trajectory of metazoan IRG1 provide insights how
305 molecular adaptations might regulate itaconate biochemistry for innate immunity. The frequent
306 loss of IRG1 in metazoan taxa, particularly the loss of IRG1L paralog in mammalian taxa,
307 clearly suggests itaconate biosynthesis is dispensable for the host's organismal central energy
308 metabolism. Due to potentially detrimental effects of itaconate in inhibiting host energy
309 metabolism, itaconate production might reduce host fitness and may therefore be subject to
310 negative selection and/or stringent regulatory strategies in evolution, a notion prompted by
311 several observations in this study.
312
313 First, we identified that the mitochondrial targeting sequence (MTS) is frequently detected in
314 pIRG1 and IRG1L, but is lost in IRG1, the only gene copy that is retained in the eutherian
315 mammals. The loss of IRG1L in the therians coupled with the loss of MTS in IRG1 would
316 relocate the catalysis outside the mitochondria and prevent direct inhibition of the mitochondrial
317 matrix-localized energy metabolism. Despite lacking MTS, mammalian IRG1 remains associated
318 with the mitochondria with unknown mechanism[59-61]. A recent identification of a direct IRG1-

319 mediated mitochondria-vacuole interface highlighted the importance of this efficient
320 mitochondria-derived itaconate production in host defense against intracellular pathogen
321 *Salmonella*[45]. Second, despite a conserved role of itaconate in metazoan innate immunity, IRG1
322 transcription in mammals is more tightly regulated than that of the basal metazoans. While the
323 mammalian IRG1 is almost exclusively restricted to LPS-activated TLR4 and interferon
324 signaling in macrophages[60,62], the pIRG1s in oysters and in amphioxus appear to be
325 transcriptionally regulated with less stringency, including a basally expressed pIRG1 in oysters'
326 tissues and hemocytes without any stimulation.
327
328 We reported a dramatic transcriptional induction of oyster pIRG1 by poly(I:C) and the low level
329 of induction by LPS stimulation, as well as a broader, dual induction of amphioxus pIRG1 by
330 both poly(I:C) and LPS stimulation. Our report on poly(I:C)-induced oyster pIRG1 expression is
331 supported by the transcriptomics analysis in mussel hemocytes upon screening for a panel of
332 pathogen-associated molecular patterns (PAMPs), where mussel pIRG1 induction is restricted to
333 poly(I:C) stimulation[63]. Lophotrochozoa (including marine bivalves) and amphioxus contain
334 rudimentary TLR sequences[64-67], however, it remains unknown if these rudimentary TLRs
335 function as pattern recognition receptors (PRRs) with the same degree of specificity and
336 downstream signaling as their mammalian counterparts[67,68]. Notably, the first identified
337 Drosophila Toll-receptor does not recognize foreign PAMPs despite the involvement in host
338 resistance against fungal and bacterial infection[69]. The strikingly high upregulation of pIRG1
339 from Poly(I:C) stimulation relative to LPS leads us to speculate that pIRG1 in oysters and
340 amphioxus might be induced by a non-TLR, membrane-bound PRRs. For instance, zebrafish
341 IRG1 is regulated by cooperation between glucocorticoid receptors and the JAK/STAT

342 pathway[20]. Future multispecies analysis on the immune signaling pathways and transcriptional
343 factors that induce IRG1 expression in non-traditional model organisms is vital to answer these
344 questions.
345
346 Investigation of three metazoan subbranches, pIRG1, IRG1L and IRG1, support the notion that
347 genome duplication, horizontal gene transfer and gene loss provide powerful evolutionary forces
348 to innovate gene function and regulation[70]. Our analysis is consistent with the notion that whole
349 genome duplication (WGD) in the vertebrate lineage that produced the paralogous IRG1L and
350 IRG1, in which IRG1L is subsequently lost in viviparous mammals. The timings of the two
351 rounds (2R) WGD of the early vertebrate lineage have been reported before[71] and after[41] the
352 gnathostome-agnathan divergence occurring around 465-428 Ma. Our findings that lamprey has
353 two pIRG1s without either IRG1 and IRG1L sequence would agree with 2R WGD and the one
354 after the divergence of agnathans and gnathostomes gave rise to IRG1 and IRG1L. The relative
355 similarity of IRG1L and pIRG1 synteny and the presence of MTS would support the notion that
356 the genome duplication and loss of IRG1L enable a tightened regulation of itaconate
357 biosynthesis exerted by extra-mitochondrial IRG1 that is subsequently retained in eutherian
358 mammals.
359
360 Our exploration of the molecular evolution of itaconate biosynthesis presents another case of
361 horizontal gene transfer in eukaryotes from prokaryotic source in the context of innate immunity.
362 The potential independent acquisition of the metazoan IRG1 and the fungi *A. terreus* ACOD
363 suggests at least one, and perhaps two horizontal gene transfer (HGT) events from bacteria to
364 eukaryotes in the evolutionary history of IRG1. The exclusive presence of ACOD in *Aspergillus*

taxa within fungi that are nested within other prokaryotic sequences would strongly suggest a HGT in *Aspergillus*. Our studies of the metazoan IRG1 can be explained by two scenarios: first, the metazoan IRG1 is derived from the last common ancestor of eukaryotes and subsequently lost several times in major eukaryotic taxa including plants, most fungi, and ecdysozoa; and second, the metazoan ancestor's IRG1 is horizontally transferred directly from bacteria. The observation that the closest branch next to metazoan IRG1 (Fig. 1C) contains sequences from alphaproteobacteria might support the former. However, the frequent local gene duplication in pIRG1 (e.g., in oysters and in sea lampreys) and IRG1L (in xenopus) suggests pIRG1/IRG1L genomic region might be highly mobile, indicative of a more recent HGT event — an analysis of mobile gene elements near pIRG1 would help resolve this question. Notably, a previous systems-level analysis of eukaryotic orthologous groups (KOGs) also highlighted CAD sequences as one of the unique eukaryotic sequences with unexpected phyletic patterns, supporting putative independent HGT events[72]. In microorganisms, the study on itaconate production in immunometabolism of other bacterial and fungi species might also indicate a universal role of itaconate in innate immunity.

Methods

Phylogenetic analysis of the IRG1 orthologs

Human IRG1 was manually queried against the NCBI database using BLASTp and tBLASTn[73] across the tree of life and were subsampled by taxonomic order during May-July 2020. Sequences from metazoan taxa were restricted to Expectation values below 1e-10. To maximize sequence diversity in more distantly related sequences from fungi, archaea, and prokaryotes inclusion criteria was relaxed to 0.001. In total, 2,297 sequences were identified. Duplicate

388 sequences were identified and removed. A full phylogeny of all sequences was constructed using
389 MAFFT (v7.471)[74]. IQTree (v2.1.2)[75] and ModelFinder[76] were used to find the best fit
390 substitution model and branch support values were obtained by ultrafast bootstrapping[77]. Figtree
391 (v1.4.4) and iTOL (v6.5.2)[78] were used for tree visualization.
392
393 **Analysis of dubious sequences**
394 Initial sequence surveys revealed only nine orthologs in Arthropoda, all in *Drosophila* genera,
395 three orthologs in Plantae: one in *Ziziphus* and two in *Quercus* generas. Reciprocal BLASTp and
396 phylogenetic analysis indicated likely sequence contamination by *Acetobacter aceti* in
397 Drosophila sequences, a black yeast-like fungal pathogen in the *Quercus* sequence, and *Belnapia*
398 contamination in the *Ziziphus* sequences. Therefore, these sequences were removed.
399
400 **Identification of mitochondrial targeting sequences (MTS)**
401 All the metazoan IRG1 sequences were analyzed using Target-P (v2.0)[79] to predict the presence
402 of N-terminal MTS. Default settings were used, and a likelihood >0.5 was considered a predicted
403 MTS.
404
405 **Synteny analysis**
406 Representative genes from the model organisms from each major metazoan taxa possessing an
407 IRG1 ortholog were manually investigated for syntenic genes using NCBI Genome Data
408 Viewer[80]. The five nearest annotated genes both upstream and downstream were analyzed using
409 Pfam[81] and NCBI Domain Architecture Retrieval Tool[82] to predict domain architecture. In

410 instances where local duplication is present, more than the five nearest genes were recorded as

411 there were multiple nearby IRG1 homologs (e.g., *Xenopus laevis*).

412

413 **Plasmids and cloning**

414 A N-terminal His-tagged *Danio rerio* IRG1L sequence lacking the N-terminus predicted

415 mitochondrial targeting sequence (Δ31aa) was codon-optimized for bacterial expression, custom

416 synthesized and subcloned into pET-28a(+)-TEV vector using Nde1 and BamH1 sites

417 (GenScript).

418

419 **His-Zebrafish IRG1L (aa 32- 466)**

420 MGSSHHHHHHSSGENLYFQGHM

421 APEETVTSSFGRFIQSVQPKNLTPTVLQRSKRMVLDSIGVGLVGSTTEVFDLALQHCQH

422 MYASDDISFVYGRQGVKLSPTLAAFVNGVAAHSMDFDDTWHPATHPSGAVLPALLAIS

423 DMLHSNSKPSGLDFLTAFNVGIEIQGRLMRFSNEAHNIPKRFHPPSVVGTLGSAAACARL

424 LSLDRNQASNALAIAASLAGAPMANAATQSKPLHIGNASRLGLEAALLASRGLEASPLV

425 LDAVPGVAGFSAFYEDYAPRPIGSPEEDEHSFLIESQDIAFKRFPAHLGMHWIADAASVV

426 HKTLVGLKGGSISPNLVQDILLRVPLSKYINRPFPESEHQARHSFQFNACTALLDGEVTV

427 QSFHPEAIQRPELHALLSRVRLEHPSDNPANFNIMYGEVEVTLVTGDVLRGRCDTFYGH

428 WRNPLSQESLHKKFRNNAGTVLSTEKVERLIEAVESLDRSDDCKQLLAQLQ*

429

430 gcgccggaggaaaccgtgaccagcagcttcggtcgttttatccagagcgttcaaccgaagaacctgaccccgaccgtgctgcagcgtagc

431 aaacgtatggttctggacagcattggtgtgggcctggttggcagcaccaccgaggtgttcgatctggcgctgcagcactgccaacacatgta

432 cgcgagcgacgatatcagctttgtgtatggtcgtcaaggcgttaagctgagcccgaccctggcggcgtttgtgaacggtgttgcggcgcac

433 agcatggactttgatgatacctggcatccggcgacccacccgagcggtgcggtgctgccggcgctgctggcgatcagcgacatgctgcac

434 agcaacagcaagccgagcggtctggatttcctgaccgcgtttaacgttggtatcgagattcagggccgtctgatgcgtttcagcaacgaagc

435 gcacaacattccgaaacgttttcacccgccgagcgtggttggtaccctgggcagcgcggcggcgtgcgcgcgtctgctgagcctggatcg

436 taaccaggcgagcaacgcgctggcgattgcggcgagcctggcgggtgcgccgatggcgaacgcggcgacccaaagcaaaccgctgc

437 acattggtaacgcgagccgtctgggtctggaggcggcgctgctggcgagccgtggtctggaagcgagcccgctggtgctggatgcggtg

438 ccgggtgttgcgggtttcagcgcgttttacgaggattatgcgccgcgtccgattggcagcccggaggaagacgagcacagcttcctgatcg

439 aaagccaggatattgcgttcaagcgttttccggcgcacctgggtatgcactggattgcggacgcggcgagcgtggttcacaagaccctggt

440 tggcctgaaaggtggcagcatcagcccgaacctggtgcaagatattctgctgcgtgttccgctgagcaaatacatcaaccgtccgttcccgg

441 agagcgaacaccaggcgcgtcacagcttccaatttaacgcgtgcaccgcgctgctggatggtgaggtgaccgttcagagctttcacccgg

442 aggcgatccaacgtccggaactgcacgcgctgctgagccgtgtgcgtctggaacacccgagcgataacccggcgaacttcaacattatgt

443 acggcgaggtggaagttaccctggtgaccggtgacgttctgcgtggccgttgcgataccttctatggtcactggcgtaacccgctgagcca

444 ggaaagcctgcacaagaaatttcgtaacaacgcgggcaccgtgctgagcaccgagaaggttgaacgtctgattgaggcggttgaaagcct

445 ggaccgtagcgacgattgcaaacaactgctggcgcagctgcaataa

446

447 For mitochondrial localization assay in the HEK293 cells, sequences were codon-optimized for

448 mammalian expression, custom synthesized and subcloned into in lentiviral pLys1 vector

449 (Addgene #50058) using Nhe1/Xma1or EcoRV/Xma1 (for MTS-human IRG1) sites (GenScript).

450

451 **Human IRG1/ACOD1-FLAG**

452 MMLKSITESFATAIHGLKVGHLTDRVIQRSKRMILDTLGAGFLGTTTEVFHIASQYSKIYS

453 SNISSTVWGQPDIRLPPTYAAFVNGVAIHSMDFDDTWHPATHPSGAVLPVLTALAEALP

454 RSPKFSGLDLLLAFNVGIEVQGRLLHFAKEANDMPKRFHPPSVVGTLGSAAAASKFLGL

455 SSTKCREALAIAVSHAGAPMANAATQTKPLHIGNAAKHGIEAAFLAMLGLQGNKQVLD

456 LEAGFGAFYANYSPKVLPSIASYSWLLDQQDVAFKRFPAHLSTHWVADAAASVRKHLV

457 AERALLPTDYIKRIVLRIPNVQYVNRPFPVSEHEARHSFQYVACAMLLDGGITVPSFHEC

458 QINRPQVRELLSKVELEYPPDNLPSFNILYCEISVTLKDGATFTDRSDTFYGHWRKPLSQE

459 DLEEKFRANASKMLSWDTVESLIKIVKNLEDLEDCSVLTTLLKGPSPPEVASNSPACNNS

460 ITNLS

461 PGGGSGGSGGSMDYKDDDDK*

462

463 gctagcgccaccatgatgctgaaaagcatcacagagagcttcgcgaccgccattcacggactgaaggtgggccacctgacagatagagt

464 gatccaaagatctaaaaggatgatcctggatacactgggcgctggattcctgggtaccaccaccgaggtgttccatatcgccagccagtaca

465 gcaagatctacagcagcaacatctctagcacagtgtggggacagcctgacatcagactgccacctacctacgccgccttcgtgaacggcgt

466 ggccatccacagcatggacttcgacgacacatggcaccccgccacccacccttccggagccgtgctgcctgtgttaacagccctggctga

467 agccctgccaagatcccctaagttcagcggcctggacctgctgctggccttcaacgtgggcatcgaggtgcaaggcagactgctgcatttc

468 gccaaggaagccaacgacatgcctaagcggttccaccctcctagcgtggtgggcacactgggcagcgccgcggccgcgtccaagtttct

469 gggcctgagcagcaccaagtgcagagaggctctggccatcgccgtgtctcacgccggcgcccctatggccaatgccgccacgcagacc

470 aagcccctgcacatcggcaatgccgccaagcacggcatcgaggccgcctttctggccatgctgggcctgcagggcaacaagcaggtgct

471 cgacctggaagccggcttcggcgccttctacgccaattacagcccaaaagtgctgcctagcatcgcctcatattcttggctgctggatcagc

472 aggacgtggccttcaagcggtttccagcccacctgagcacccactgggtggctgatgccgctgccagcgtcagaaagcacctggtggca

473 gaaagagccctgctgcccaccgactacatcaaacggattgtgctgagaatacccaacgtgcagtacgtgaaccggcccttccctgtcagcg

474 agcacgaggctagacactcttttcagtatgtggcttgtgccatgctactggacggcggaatcaccgtgccttcttttcacgagtgccagatcaa

475 cagacctcaggtccgcgaactgctgtccaaggtggaactcgaatacccccccgataacctgcctagcttcaacatcctgtactgcgagatca

476 gcgtgaccctgaaggacggcgccacattcaccgacagaagcgacaccttctacggccactggcggaagcctctgtctcaggaggatctg

477 gaggaaaagttcagagctaacgcctccaaaatgctgagctgggacaccgtggaaagcctgatcaagatcgttaagaacctggaggacctg

gaggactgcagcgtgttgacaaccctgctgaagggcccttcccctcccgaggttgctagtaatagccctgcctgtaacaacagcatcacca

acctgtcccccgggggtggatctggtggatctggtggatctatggattacaaggatgacgatgacaagtaa

MTS-human IRG1L-FLAG (with MTS from cytochrome C oxidase subunit 4I1 (COX4I1))

MLATRVFSLVGKRAISTSVCVRAHAS

MMLKSITESFATAIHGLKVGHLTDRVIQRSKRMILDTLGAGFLGTTTEVFHIASQYSKIYS

SNISSTVWGQPDIRLPPTYAAFVNGVAIHSMDFDDTWHPATHPSGAVLPVLTALAEALP

RSPKFSGLDLLLAFNVGIEVQGRLLHFAKEANDMPKRFHPPSVVGTLGSAAAASKFLGL

SSTKCREALAIAVSHAGAPMANAATQTKPLHIGNAAKHGIEAAFLAMLGLQGNKQVLD

LEAGFGAFYANYSPKVLPSIASYSWLLDQQDVAFKRFPAHLSTHWVADAAASVRKHLV

AERALLPTDYIKRIVLRIPNVQYVNRPFPVSEHEARHSFQYVACAMLLDGGITVPSFHEC

QINRPQVRELLSKVELEYPPDNLPSFNILYCEISVTLKDGATFTDRSDTFYGHWRKPLSQE

DLEEKFRANASKMLSWDTVESLIKIVKNLEDLEDCSVLTTLLKGPSPPEVASNSPACNNS

ITNLS

PGGGSGGSGGSMDYKDDDDK*

gatatcgccaccatgctggctacaagagtgttcagcctggtcggcaagagagccatcagcacctctgtgtgcgtgcgggcccacgctagc

atgatgctgaaaagcatcacagagagcttcgcgaccgccattcacggactgaaggtgggccacctgacagatagagtgatccaaagatct

aaaaggatgatcctggatacactgggcgctggattcctgggtaccaccaccgaggtgttccatatcgccagccagtacagcaagatctaca

gcagcaacatctctagcacagtgtggggacagcctgacatcagactgccacctacctacgccgccttcgtgaacggcgtggccatccaca

gcatggacttcgacgacacatggcaccccgccacccacccttccggagccgtgctgcctgtgttaacagccctggctgaagccctgccaa

gatcccctaagttcagcggcctggacctgctgctggccttcaacgtgggcatcgaggtgcaaggcagactgctgcatttcgccaaggaagc

caacgacatgcctaagcggttccaccctcctagcgtggtgggcacactgggcagcgccgcggccgcgtccaagtttctgggcctgagca

501 gcaccaagtgcagagaggctctggccatcgccgtgtctcacgccggcgcccctatggccaatgccgccacgcagaccaagcccctgca

502 catcggcaatgccgccaagcacggcatcgaggccgcctttctggccatgctgggcctgcagggcaacaagcaggtgctcgacctggaag

503 ccggcttcggcgccttctacgccaattacagcccaaaagtgctgcctagcatcgcctcatattcttggctgctggatcagcaggacgtggcct

504 tcaagcggtttccagcccacctgagcacccactgggtggctgatgccgctgccagcgtcagaaagcacctggtggcagaaagagccctg

505 ctgcccaccgactacatcaaacggattgtgctgagaatacccaacgtgcagtacgtgaaccggcccttccctgtcagcgagcacgaggcta

506 gacactcttttcagtatgtggcttgtgccatgctactggacggcggaatcaccgtgccttcttttcacgagtgccagatcaacagacctcaggt

507 ccgcgaactgctgtccaaggtggaactcgaatacccccccgataacctgcctagcttcaacatcctgtactgcgagatcagcgtgaccctga

508 aggacggcgccacattcaccgacagaagcgacaccttctacggccactggcggaagcctctgtctcaggaggatctggaggaaaagttc

509 agagctaacgcctccaaaatgctgagctgggacaccgtggaaagcctgatcaagatcgttaagaacctggaggacctggaggactgcag

510 cgtgttgacaaccctgctgaagggcccttcccctcccgaggttgctagtaatagccctgcctgtaacaacagcatcaccaacctgtcccccg

511 gg

512 **zebrafish IRG1L-FLAG**

513 MLSAVQRSSRYLTSFSAARGLHKSALDVAERPAPEETVTSSFGRFIQSVQPKNLTPTVLQ

514 RSKRMVLDSIGVGLVGSTTEVFDLALQHCQHMYASDDISFVYGRQGVKLSPTLAAFVN

515 GVAAHSMDFDDTWHPATHPSGAVLPALLAISDMLHSNSKPSGLDFLTAFNVGIEIQGRL

516 MRFSNEAHNIPKRFHPPSVVGTLGSAAACARLLSLDRNQASNALAIAASLAGAPMANA

517 ATQSKPLHIGNASRLGLEAALLASRGLEASPLVLDAVPGVAGFSAFYEDYAPRPIGSPEE

518 DEHSFLIESQDIAFKRFPAHLGMHWIADAASVVHKTLVGLKGGSISPNLVQDILLRVPLS

519 KYINRPFPESEHQARHSFQFNACTALLDGEVTVQSFHPEAIQRPELHALLSRVRLEHPSD

520 NPANFNIMYGEVEVTLVTGDVLRGRCDTFYGHWRNPLSQESLHKKFRNNAGTVLSTEK

521 VERLIEAVESLDRSDDCKQLLAQLQ

522 PGGGSGGSGGSMDYKDDDDK*

523

524 gctagcgccaccatgctgtctgccgttcagaggtccagcagatacctgaccagcttcagcgccgctagaggcctccacaaaagcgccctg

525 gacgtggccgagaggcctgctcctgaagagacagtgaccagcagcttcggcagattcatccagtctgtgcaacctaagaacctgacacct

526 accgtgctgcaacgctccaaaagaatggtgctcgactccatcggcgtgggcctggtgggcagcacaaccgaggtgttcgacttggccctg

527 cagcactgccagcatatgtatgcctccgacgacatcagctttgtgtacggccggcagggcgtgaagctgagcccaacactggctgctttcgt

528 gaacggcgtggccgctcatagcatggacttcgacgatacctggcaccctgccacccacccttcaggcgccgtgctccctgctctgctggct

529 atcagcgacatgctgcattctaatagcaagccaagcgggctggacttcctgaccgcctttaacgtgggcatcgaaatccagggaagactga

530 tgagatttagcaacgaggcccacaacatccccaagagattccacccacctagcgtggtgggcaccctgggctctgctgcagcttgtgccag

531 actgctgagcctggacagaaaccaggccagcaacgccctggccattgccgccagtctggccggcgcccctatggccaacgccgccaca

532 cagagtaagcctctgcacatcggcaacgccagccggctgggcctggaagccgccctgctggccagcagaggcctggaagccagccctc

533 tggtcctggatgccgtgcccggcgtcgccggatttagcgccttctacgaggactacgcccctagacctatcggatctcctgaggaagatga

534 gcacagcttcctgatcgagagccaggacatcgccttcaagcggttccccgcccacctgggcatgcactggatcgccgacgccgccagcg

535 tggtgcacaagaccctggtgggcctgaagggcggatctatcagccccaatctggtgcaggacatcctgctgcgggtgcctctgtccaagta

536 catcaacagacccttccctgagagcgagcaccaggcccggcacagcttccagttcaacgcctgcaccgccctgctggatggagaggtga

537 ccgtgcagagctttcaccccgaagcaattcaaagacccgagctgcacgccctgttgtccagagtgcggctggaacacccatctgacaacc

538 ccgctaatttcaacatcatgtacggcgaggtcgaggtgaccctggtgacaggcgacgtgctgagaggacggtgcgacaccttctacggcc

539 actggcggaatcctctgtctcaggagagcctacacaaaaagttcagaaacaacgctggtacagttctgagcaccgagaaggtggaaagac

540 tgatcgaagccgtggaaagcctggatagatctgatgattgcaagcagctgctggcccagctgcagcccggg

541

542 **Zebrafish IRG1/ACOD1-FLAG**

543 MIRKSVTDSFGAAVSCLSTSHLTDEVIRRSKRMILDTLGVGLIGTRTPVFNTVLQCSQRQ

544 QALENSKVWGRPGSSLPPQYAAFVNGVAVHSMDFDDTWHPATHPSGAVLPALLALAE

545 TLPVKPSGLELLLAFNVGIEIQGRLLRFSKEAHNIPKRFHPPAVVGVMGSAAATAKLLRL

546 PAAQSIAALAIACSSAGAPMANAATQTKPLHMGNAARGGLEASQLALFGLVGNTHILDL

PSGFGSFYPDYVPHPLAEVTPNSHYRWVLEEQDIAQKRFPAHLGMHWVADAAIEARAK
FLDKHPNADLSQIKKITLRVPSSRYVDCPLPVTEHQARHSFQFNCCTALLDGEVNVQSFS
RAQMNRSVLKELLLKVELENPQDNHSSFQKMYCEITVLSTQGEMFTARCDTFYGHWRK
PLSQEDLLKKFRLNASSVLPNEVLEEIIYAVDHLDTNQDCSKLWSYMHFNKHMVQRDE
RRCFALA

PGGGSGGSGGSMDYKDDDDK

gctagcgccaccatgatccggaaaagcgtgacagatagctttggagccgccgtgtcctgcctgtctacaagccacctgaccgatgaggtg
atcagaagaagcaagcggatgatcctggacacactgggcgtgggcctgattggaacaagaacccctgtgttcaacaccgtgctgcagtgc
agccaaagacagcaggccctggaaaacagcaaggtgtggggcagacccggctcttccctgcctcctcaatacgccgccttcgtgaacgg
cgtggccgtgcacagcatggactttgacgacacctggcaccccgccacacatcctagcggtgccgtgctgcctgctctgctggccctggct
gaaaccctgccagtgaagccttctggcctggagctgctgctggcctttaatgtcggcatcgagattcagggcagactgctgagattcagcaa
agaggcccacaacatccccaagagattccacccacctgctgttgtgggagtcatgggcagcgctgccgccacagccaagctgctgcggc
tgcctgccgcccagagcatcgccgctctggctatcgcatgcagcagcgccggagcccctatggccaacgccgccacccagaccaagcc
cctgcacatgggcaacgccgctagaggcggcctggaggccagccagctggctctgttcggcctggtggggaatacccacatcctggatct
gccatctggcttcggctccttttaccccgactacgtgcctcacccccctggcagaagtgacccctaacagccattatagatgggtgcttgaaga
gcaggacatcgctcagaagcgcttccctgcccacctgggaatgcactgggtcgccgacgccgccatcgaggccagagctaagttcctgg
acaagcaccctaatgccgatctgtcgcagatcaagaaaatcacactgagagtgcccagctctagatacgtggactgtcctctgccagtgaca
gagcaccaggccagacactcttttcagttcaattgctgcaccgccctgctggacggcgaagtgaacgtgcaaagcttcagcagagcccag
atgaaccggagcgtgctgaaggaactgctgctcaaagtggaactggaaaacccccaggacaaccacagctccttccagaaaatgtactgc
gagatcaccgtgctgagcacccagggcgagatgttcaccgccaggtgtgataccttctacggccactggcggaagcctctgtctcaggag
gacctgcttaagaagttccggctgaacgccagcagcgtgctccctaacgaggtcctggaagagatcatctacgccgtggaccacctggata

569 ccaaccaggattgttccaagctgtggtcctacatgcatttcaacaagcacatggtgcagcgggacgagcggagatgcttcgccctggcccc
570 cggg
571
572 **Submitochondrial localization by limited proteolysis**
573 Raw mitochondrial fraction from HEK293 cells stably expressing zebrafish IRG1L-FLAG or
574 MTS-human IRG1/ACOD1-FLAG were subjected to limited proteolysis following the previous
575 protocol[83]. Specifically, 30 million cells were washed with cold PBS, scraped into 1 ml ice cold
576 mito isolation buffer (10 mM Tris-HCl, 1 mM EGTA-Tris, 200 mM sucrose, pH 7.4) that also
577 contains sub-optimal concentration of protease inhibitor (cOmplete™, Mini, EDTA-free Protease
578 Inhibitor Cocktail, Roche) at 1 mini tablet in 30 ml isolation buffer, and homogenized with 20
579 strokes in 2 ml douncer homogenizer (KIMBLE). The homogenates were spun at 600 g for 10
580 min at 4 °C, the supernatants were spun at 7000 g for 10 min at 4 °C, and the mitos in the pellets
581 were resuspended in 250 µl mito isolation buffer with no protease inhibitor. The mito
582 homogenates were spun again at 600 g for 3 min at 4 °C to remove nucleus, and aliquoted into
583 PCR tubes at approximately 20 µg protein concentration. The mitos were treated with 100 µg/ml
584 proteinase K (Millipore Sigma, P2308) in the presence of increasing concentrations of digitonin
585 (Millipore Sigma, BN2006) at 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2 and 0.4% digitonin, or 1%
586 Triton X-100 for 15 min at room temperature. 7 mM PMSF was added to inactivate proteinase K
587 for 5 min prior to SDS-PAGE analysis and western blotting.
588
589 **Zebrafish IRG1L protein production**
590 The His-Zebrafish IRG1L plasmid was transformed in BL21 Star™ (DE3) Chemically
591 Competent E. coli (Invitrogen c6010-03), grown at 225 rpm at 37 °C to OD600 0.6. Cells were

592 then induced with 0.1 mM β-D-1-thiogalactopyranoside (IPTG) for 6 hours at 18 °C. Cells were
593 lysed by sonication (Qsonica Q500 with microtip) in 200 mL lysis buffer (20mM HEPES pH
594 7.4, 500 mM NaCl, 10% glycerol v/v, 1 mM TCEP (pH 7.4), 100 U/mL DNase, 0.1 mg/mL
595 Lysozyme, and cOmplete mini protease inhibitor cocktail) using 2 sec on and 2 sec off pulses at
596 25% power for 2 minutes. The recombinant IRG1-like protein was enriched using cobalt
597 TALON Metal Affinity Resin (Takara Bio #635501) with gentle shaking for 1 hour at 4 °C and
598 washed twice by centrifugation for 3 min at 3500 rpm at 4 °C and resuspension in the washing
599 buffer (20 mM Tris pH 8.0, 150 mM NaCl, 20 mM Imidazole, 1 mM TCEP). Protein was eluted
600 from the affinity resin in the elution buffer (20 mM Tris pH 8.0, 150 mM NaCl, 500 mM
601 Imidazole, 1 mM TCEP, 10% glycerol). The recombinant protein was then concentrated using
602 centrifugal filter units (Amicon Ultra-4, Millipore) and dialyzed using Slide-A-Lyzer dialysis
603 cassette into storage buffer (20 mM Tris pH 8.0, 150 mM NaCl, 20% glycerol, 1 mM TCEP)
604 overnight.
605
606 **Enzymology**
607 The enzyme assay for the zebrafish IRG1L was performed by incubating 7 µg recombinant
608 protein with 1 mM cis-aconitate (Sigma A3412-1G) in 300 µl reaction buffer (10 mM HEPES,
609 150 mM NaCl, 0.1 mM TCEP, 10% glycerol, pH 6.5) at 37 °C. Aliquots of 50 µl were taken at
610 0, 10, 30, and 60 minutes and were immediately quenched with 2 % formic acid, vortexed, and
611 centrifuged at 14,000 rpm for 20 minutes at 4 °C. The supernatant was then transferred to a new
612 centrifuge tube and mixed 1:9 with a 3:1 acetonitrile:methanol solvent for LC-MS analysis
613 following the established polar metabolite profiling protocol in the lab[84]. Itaconate production
614 was confirmed using chemical standard itaconic acid (Sigma Aldrich #I29204).

615

616 **Mollusk pIRG1 expression**

617 Eastern oysters (*Crassostrea virginica*, Woods Hole Marine Biological Laboratory) were
618 acclimated in filtered sea water overnight after arrival. To determine unstimulated basal pIRG1
619 expression, oysters (n = 5) were dissected on ice and 50 mg wet tissue mass from the gill,
620 mantle, stomach, adductor muscle, and heart were collected. Tissue was washed in PBS, then
621 homogenized in RL+BME buffer (Qiagen). mRNA extraction was performed per manufacturer's
622 instruction using the RNeasy mini kit (Qiagen). The mRNA concentration was determined using
623 a Nanodrop and cDNA libraries were prepared using SuperScriptIII First-Strand Synthesis Super
624 Mix for qRT-PCR (Thermo Fisher Scientific).

625

626 **Mollusk genes quantitative PCR**

627 The control gene EF1a for qPCR was identified from the literature[85]. Primers for EF1a and
628 pIRG1 were designed using primer-BLAST (NCBI) and synthesized from Integrated DNA
629 technologies. The pilot qPCR and Sanger sequencing experiments were conducted to validate the
630 primers for pIRG1 and EF1a. qPCR was performed using PowerUp SYBR Green Master Mix
631 (Thermo Fisher A25742) in a 96-well format with Thermo Fisher QuantStudio 6Flex instrument.

Organism	Gene	Accession Number	Direction	Primer Sequence
Eastern Oyster (*C. virginica*)	IRG1	XM_022457065.1	F	5' - TGG TAT CTC GGC AAA CGG AC - 3'
Eastern Oyster (*C. virginica*)	IRG1	XM_022457065.1	R	5' - TTC GTT GAA ACG GCT CTC ACT - 3'
Eastern Oyster (*C. virginica*)	EF1-a	XM_022472315	F	5' - CAT CCC GCT ACC CAT CCA AG - 3'
Eastern Oyster (*C. virginica*)	EF1-a	XM_022472315	R	5' - CGC TGG TGG ATG GAG TCT ATT - 3'

632

633

634 **Mollusk Hemocyte Assays**

635 For all hemocyte assays, live eastern oysters were notched and hemolymph extracted by syringe
636 from the adductor muscles following published protocol[46]. Hemolymph from 10 oysters was
637 pooled on ice, diluted 1:2 with decontamination washing buffer (PBS, 100 µg/mL penicillin-
638 streptomycin), pipetted into Corning Falcon 60 mm tissue culture plates (Thermo Fisher) and
639 were incubated in the dark at room temperature for 30 min. Cells were then inspected under the
640 microscope to ensure adhesion, and decontamination buffer was aspirated and replaced with 5
641 mL culture media (1:1 OptiMEM media: filtered sea water, 50 µg/mL penicillin-streptomycin).
642

643 The cultured hemocytes were treated with zymosan (100 µg/mL, Invivogen tlrl-zyd), heat-killed
644 *E. coli* ($1x10^7$ PFU, Invivogen tlrl-hkeb2), heat-killed *S. aureus* ($1x10^7$ PFU, Invivogen tlrl-
645 hksa), and Poly(I:C) (10 µg/mL final concentration, Invivogen tlrl-pic), and filtered sea water
646 (control) for 8 hrs in the dark at room temperature approximately 20 °C, or Poly(I:C) and LPS E.
647 coli 0111:B4 (1 µg/mL final concentration, Sigma LPS25) for indicated time. After treatments,
648 cells were directly lysed and analyzed as described above.
649

650 **Amphioxus pIRG1 expression**

651 Amphioxus (*Branchiostoma floridae*, Gulf Specimen Marine Lab) were kept in the filtered sea
652 water in the dark at room temperature. Three amphioxus each condition were treated by oral
653 injection with LPS E. coli 0111:B4 (50 µg/mL final concentration, Sigma LPS25), Poly(I:C)(1
654 mg/mL, Invivogen tlrl-pic), or control (filtered seawater). Specifically, amphioxus were cold-
655 immobilized and a 25 gauge needle was used to carefully administer 100 µl stimulus solution
656 through the oral hood. Individuals were immediately placed back into their containers at room

657 temperature and recovered overnight. Individuals were netted, cold immobilized and dissected
658 24 hr post treatment. Gill, muscle, hepatocecum and intestine tissue were collected and washed
659 in PBS before tissue lysis and mRNA expression analysis.
660
661 Candidate reference genes were identified from literature[86]. Primers for pIRG1 and reference
662 genes S20 and EF1a were designed using NCBI primer-BLAST, and synthesized (Integrated
663 DNA technologies):

Organism	Gene	Accession Number	Direction	Primer Sequence
Amphioxus (*B. floridae*)	IRG1	XM_035821914.1	F	5' - ATC TTA GAC ATG CGG AGC GG - 3'
Amphioxus (*B. floridae*)	IRG1	XM_035821914.1	R	5' - TTG ATC GCC ACG TCC TGT TT - 3'
Amphioxus (*B. floridae*)	EF1-a	XM_035826544.1	F	5' - CGG CGG GAA GAA GAA GTG AT - 3'
Amphioxus (*B. floridae*)	EF1-a	XM_035826544.1	R	5' - AGA ACG GGT CGG TAA CGA TT - 3'
Amphioxus (*B. floridae*)	S20	XM_035811399.1	F	5' - TCC GCA TGC CTA CAA AGG TC - 3'
Amphioxus (*B. floridae*)	S20	XM_035811399.2	R	5' - TGC TGA TGG ATG TGA TCT GC - 3'

664
665
666 **Statistics and reproducibility**
667 For gene expression experiments, data are shown as mean ± s.d., $n \geq 3$ biologically independent
668 samples unless stated otherwise. Statistical significance was calculated using two-tailed t test.
669 Significance level were indicated as **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ and
670 n.s. $p > 0.05$. Statistical analysis was performed using GraphPad Prism 9.3.1, or as reported by
671 the relevant computational tools.
672
673 **Data availability**

674 The data that support this study are available from the corresponding author upon reasonable
675 request. The sequences used to for phylogenetic inference (.FASTA files) are available within
676 the supplementary data.
677
678 **Author Contributions**
679 H.S. and G.W. conceived the initial idea. R.S., A.S., and M.D. performed bioinformatics
680 analysis; H.S., R.S., and F.M. performed recombinant protein production and enzymatic assays;
681 R.S. and H.S. performed dissections and qRT-PCR assays. H.S. and R.S. wrote the initial
682 manuscript with input from G.W. All authors reviewed and approved the manuscript.
683
684 **Acknowledgements**
685 We thank Yale School of Medicine, Yale West Campus, and Yale Systems Biology Institute for
686 the instrumentation support; X. Shi for technical support on the LC-MS; J. Liu for technical
687 support; A. Chavan and L. Wu on advise on assays in amphioxus and oyster; K. Byrne for her
688 insight in sequence analysis; and members of the Shen lab, D. Stadtmauer in the Wagner lab, M.
689 Murphy, J. MacMicking, A. Goodman and J. Galán for fruitful discussion and feedback. The
690 work is supported by Yale School of Medicine startup fund and Yale Systems Biology Institute
691 pilot award, John Templeton Foundation to G.W. (#61329), Klingenstein-Simons fellowship
692 Awards in Neuroscience, 1907 Foundation Trailblazer Award, and Rita Allen Foundation to
693 H.S. and a National Science Foundation Graduate Research Fellowship to R.S. (# DGE-
694 2036201).
695
696 **References:**

1. Wu, J. & Chen, Z. J. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**, 461-488 (2014).
2. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat Microbiol* **5**, 1608-1615 (2020).
3. Duncan-Lowey, B. & Kranzusch, P. J. CBASS phage defense and evolution of antiviral nucleotide signaling. *Curr Opin Immunol* **74**, 156-163 (2022).
4. Wu, X. *et al*. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res* **42**, 8243-8257 (2014).
5. Chou, S. *et al*. Transferred interbacterial antagonism genes augment eukaryotic innate immune function. *Nature* **518**, 98-101 (2015).
6. Strelko, C. L. *et al*. Itaconic acid is a mammalian metabolite induced during macrophage activation. *J Am Chem Soc* **133**, 16386-16389 (2011).
7. O'Neill, L. A. J. & Artyomov, M. N. Itaconate: the poster child of metabolic reprogramming in macrophage function. *Nat Rev Immunol* **19**, 273-281 (2019).
8. Hooftman, A. & O'Neill, L. A. J. The Immunomodulatory Potential of the Metabolite Itaconate. *Trends Immunol* **40**, 687-698 (2019).
9. Michelucci, A. *et al*. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc Natl Acad Sci U S A* **110**, 7820-7825 (2013).
10. McFadden, B. A. & Purohit, S. Itaconate, an isocitrate lyase-directed inhibitor in Pseudomonas indigofera. *J Bacteriol* **131**, 136-144 (1977).
11. Patel, T. R. & McFadden, B. A. Caenorhabditis elegans and Ascaris suum: inhibition of isocitrate lyase by itaconate. *Exp Parasitol* **44**, 262-268 (1978).
12. Shen, H. *et al*. The Human Knockout Gene CLYBL Connects Itaconate to Vitamin B12. *Cell* **171**, 771-782 e711 (2017).
13. Ruetz, M. *et al*. Itaconyl-CoA forms a stable biradical in methylmalonyl-CoA mutase and derails its activity and repair. *Science* **366**, 589-593 (2019).
14. Nemeth, B. *et al*. Abolition of mitochondrial substrate-level phosphorylation by itaconic acid produced by LPS-induced Irg1 expression in cells of murine macrophage lineage. *FASEB J* **30**, 286-300 (2016).
15. Cordes, T. *et al*. Immunoresponsive Gene 1 and Itaconate Inhibit Succinate Dehydrogenase to Modulate Intracellular Succinate Levels. *J Biol Chem* **291**, 14274-14284 (2016).
16. Lampropoulou, V. *et al*. Itaconate Links Inhibition of Succinate Dehydrogenase with Macrophage Metabolic Remodeling and Regulation of Inflammation. *Cell Metab* **24**, 158-166 (2016).
17. Mills, E. L. *et al*. Itaconate is an anti-inflammatory metabolite that activates Nrf2 via alkylation of KEAP1. *Nature* **556**, 113-117 (2018).
18. Bambouskova, M. *et al*. Electrophilic properties of itaconate and derivatives regulate the IkappaBzeta-ATF3 inflammatory axis. *Nature* **556**, 501-504 (2018).
19. Sasikaran, J., Ziemski, M., Zadora, P. K., Fleig, A. & Berg, I. A. Bacterial itaconate degradation promotes pathogenicity. *Nat Chem Biol* **10**, 371-377 (2014).
20. Hall, C. J. *et al*. Immunoresponsive gene 1 augments bactericidal activity of macrophage-lineage cells by regulating beta-oxidation-dependent mitochondrial ROS production. *Cell Metab* **18**, 265-278 (2013).
21. Hall, C. J. *et al*. Epidermal cells help coordinate leukocyte migration during inflammation through fatty acid-fuelled matrix metalloproteinase production. *Nat Commun* **5**, 3880 (2014).
22. Daniels, B. P. *et al*. The Nucleotide Sensor ZBP1 and Kinase RIPK3 Induce the Enzyme IRG1 to Promote an Antiviral Metabolic State in Neurons. *Immunity* **50**, 64-76 e64 (2019).
23. Kanamasa, S., Dwiarti, L., Okabe, M. & Park, E. Y. Cloning and functional characterization of the cis-aconitic acid decarboxylase (CAD) gene from Aspergillus terreus. *Appl Microbiol Biotechnol* **80**, 223-229 (2008).
24. Li, A. *et al*. A clone-based transcriptomics approach for the identification of genes relevant for itaconic acid production in Aspergillus. *Fungal Genet Biol* **48**, 602-611 (2011).

25. Geiser, E. *et al*. Ustilago maydis produces itaconic acid via the unusual intermediate trans-aconitate. *Microb Biotechnol* **9**, 116-126 (2016).
26. Chen, F. *et al*. Crystal structure of cis-aconitate decarboxylase reveals the impact of naturally occurring human mutations on itaconate synthesis. *Proc Natl Acad Sci U S A* **116**, 20644-20654 (2019).
27. Dolan, S. K. & Welch, M. The Glyoxylate Shunt, 60 Years On. *Annu Rev Microbiol* **72**, 309-330 (2018).
28. Wierckx, N. *et al*. Metabolic specialization in itaconic acid production: a tale of two fungi. *Curr Opin Biotechnol* **62**, 153-159 (2020).
29. Galperin, M. Y. & Koonin, E. V. Divergence and convergence in enzyme evolution. *J Biol Chem* **287**, 21-28 (2012).
30. Brock, M., Maerker, C., Schutz, A., Volker, U. & Buckel, W. Oxidation of propionate to pyruvate in Escherichia coli. Involvement of methylcitrate dehydratase and aconitase. *Eur J Biochem* **269**, 6184-6194 (2002).
31. Horswill, A. R. & Escalante-Semerena, J. C. Salmonella typhimurium LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J Bacteriol* **181**, 5615-5623 (1999).
32. Bramer, C. O. & Steinbuchel, A. The methylcitric acid pathway in Ralstonia eutropha: new genes identified involved in propionate metabolism. *Microbiology (Reading)* **147**, 2203-2214 (2001).
33. Claes, W. A., Puhler, A. & Kalinowski, J. Identification of two prpDBC gene clusters in Corynebacterium glutamicum and their involvement in propionate degradation via the 2-methylcitrate cycle. *J Bacteriol* **184**, 2728-2739 (2002).
34. Reddick, J. J. *et al*. First Biochemical Characterization of a Methylcitric Acid Cycle from Bacillus subtilis Strain 168. *Biochemistry* **56**, 5698-5711 (2017).
35. Savvi, S. *et al*. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium tuberculosis: implications for propionate metabolism during growth on fatty acids. *J Bacteriol* **190**, 3886-3895 (2008).
36. Griffin, J. E. *et al*. Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. *Chem Biol* **19**, 218-227 (2012).
37. Munoz-Elias, E. J., Upton, A. M., Cherian, J. & McKinney, J. D. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. *Mol Microbiol* **60**, 1109-1122 (2006).
38. Uchiyama, H., Ando, M., Toyonaka, Y. & Tabuchi, T. Subcellular localization of the methylcitric-acid-cycle enzymes in propionate metabolism of Yarrowia lipolytica. *Eur J Biochem* **125**, 523-527 (1982).
39. Andersson, S. G. *et al*. The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**, 133-140 (1998).
40. Simakov, O. *et al*. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol* **4**, 820-830 (2020).
41. Berthelot, C. *et al*. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* **5**, 3657 (2014).
42. Prasad, A. B., Allard, M. W., Program, N. C. S. & Green, E. D. Confirming the phylogeny of mammals by use of large comparative sequence data sets. *Mol Biol Evol* **25**, 1795-1808 (2008).
43. Omura, T. Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria. *J Biochem* **123**, 1010-1016 (1998).
44. Nutzmann, H. W., Scazzocchio, C. & Osbourn, A. Metabolic Gene Clusters in Eukaryotes. *Annu Rev Genet* **52**, 159-183 (2018).
45. Lian, H. *et al*. Parkinson's disease kinase LRRK2 coordinates a cell-intrinsic itaconate-dependent defence pathway against intracellular Salmonella. *Nat Microbiol* **8**, 1880-1895 (2023).
46. Potts, R. W. A., Gutierrez, A. P., Cortes-Araya, Y., Houston, R. D. & Bean, T. P. Developments in marine invertebrate primary culture reveal novel cell morphologies in the model bivalve Crassostrea gigas. *PeerJ* **8**, e9180 (2020).

47. Canesi, L., Gallo, G., Gavioli, M. & Pruzzo, C. Bacteria-hemocyte interactions and phagocytosis in marine bivalves. *Microsc Res Tech* **57**, 469-476 (2002).
48. Hine, P. M. The inter-relationships of bivalve haemocytes. *Fish & Shellfish Immunology* **9**, 367-385 (1999).
49. Ford, S. E., Ashtonalcox, K. A. & Kanaley, S. A. Comparative Cytometric and Microscopic Analyses of Oyster Hemocytes. *Journal of Invertebrate Pathology* **64**, 114-122 (1994).
50. Metchnikoff, E. & Binnie, F. G. *Immunity in infective diseases*. (University Press, 1905).
51. Tauber, A. I. Metchnikoff and the phagocytosis theory. *Nat Rev Mol Cell Biol* **4**, 897-901 (2003).
52. Chen, J. Y., Huang, D. Y. & Li, C. W. An early Cambrian craniate-like chordate. *Nature* **402**, 518-522 (1999).
53. Huang, S. *et al*. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res* **18**, 1112-1126 (2008).
54. Rhodes, C. P., Ratcliffe, N. A. & Rowley, A. F. Presence of coelomocytes in the primitive chordate amphioxus (Branchiostoma lanceolatum). *Science* **217**, 263-265 (1982).
55. Young, T. *et al*. Differential expression of novel metabolic and immunological biomarkers in oysters challenged with a virulent strain of OsHV-1. *Dev Comp Immunol* **73**, 229-245 (2017).
56. Sendra, M., Saco, A., Rey-Campos, M., Novoa, B. & Figueras, A. Immune-responsive gene 1 (IRG1) and dimethyl itaconate are involved in the mussel immune response. *Fish Shellfish Immunol* **106**, 645-655 (2020).
57. Nguyen, T. V., Alfaro, A. C., Young, T., Ravi, S. & Merien, F. Metabolomics Study of Immune Responses of New Zealand Greenshell Mussels (Perna canaliculus) Infected with Pathogenic Vibrio sp. *Mar Biotechnol (NY)* **20**, 396-409 (2018).
58. Nguyen, T. V. *et al*. Itaconic acid inhibits growth of a pathogenic marine Vibrio strain: A metabolomics approach. *Sci Rep* **9**, 5937 (2019).
59. Chen, M. *et al*. Itaconate is an effector of a Rab GTPase cell-autonomous host defense pathway against Salmonella. *Science* **369**, 450-455 (2020).
60. Tallam, A. *et al*. Gene Regulatory Network Inference of Immunoresponsive Gene 1 (IRG1) Identifies Interferon Regulatory Factor 1 (IRF1) as Its Transcriptional Regulator in Mammalian Macrophages. *PLoS One* **11**, e0149050 (2016).
61. Degrandi, D., Hoffmann, R., Beuter-Gunia, C. & Pfeffer, K. The proinflammatory cytokine-induced IRG1 protein associates with mitochondria. *J Interferon Cytokine Res* **29**, 55-67 (2009).
62. Naujoks, J. *et al*. IFNs Modify the Proteome of Legionella-Containing Vacuoles and Restrict Infection Via IRG1-Derived Itaconic Acid. *PLoS Pathog* **12**, e1005408 (2016).
63. Moreira, R. *et al*. Stimulation of Mytilus galloprovincialis Hemocytes With Different Immune Challenges Induces Differential Transcriptomic, miRNomic, and Functional Responses. *Front Immunol* **11**, 606102 (2020).
64. Zhang, L., Li, L. & Zhang, G. A Crassostrea gigas Toll-like receptor and comparative analysis of TLR pathway in invertebrates. *Fish Shellfish Immunol* **30**, 653-660 (2011).
65. Gerdol, M. & Venier, P. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol* **46**, 17-38 (2015).
66. Sasaki, N., Ogasawara, M., Sekiguchi, T., Kusumoto, S. & Satake, H. Toll-like receptors of the ascidian Ciona intestinalis: prototypes with hybrid functionalities of vertebrate Toll-like receptors. *J Biol Chem* **284**, 27336-27343 (2009).
67. Beutler, B. & Rietschel, E. T. Innate immune sensing and its roots: the story of endotoxin. *Nat Rev Immunol* **3**, 169-176 (2003).
68. Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nat Immunol* **5**, 987-995 (2004).
69. Lemaitre, B. & Hoffmann, J. The host defense of Drosophila melanogaster. *Annu Rev Immunol* **25**, 697-743 (2007).
70. Mottes, F., Villa, C., Osella, M. & Caselle, M. The impact of whole genome duplications on the human gene regulatory networks. *PLoS Comput Biol* **17**, e1009638 (2021).

850 71. Holland, L. Z. & Ocampo Daza, D. A new look at an old question: when did the second whole
851 genome duplication occur in vertebrate evolution? *Genome Biol* **19**, 209 (2018).
852 72. Koonin, E. V. *et al*. A comprehensive evolutionary classification of proteins encoded in complete
853 eukaryotic genomes. *Genome Biol* **5**, R7 (2004).
854 73. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search
855 tool. *J Mol Biol* **215**, 403-410 (1990).
856 74. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements
857 in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).
858 75. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective
859 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274
860 (2015).
861 76. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder:
862 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587-589 (2017).
863 77. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the
864 Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).
865 78. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display
866 and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021).
867 79. Almagro Armenteros, J. J. *et al*. Detecting sequence signals in targeting peptides using deep learning.
868 *Life Sci Alliance* **2** (2019).
869 80. Rangwala, S. H. *et al*. Accessing NCBI data using the NCBI Sequence Viewer and Genome Data
870 Viewer (GDV). *Genome Res* **31**, 159-169 (2021).
871 81. Mistry, J. *et al*. Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419
872 (2021).
873 82. Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology by domain
874 architecture. *Genome Res* **12**, 1619-1623 (2002).
875 83. Sancak, Y. *et al*. EMRE is an essential component of the mitochondrial calcium uniporter complex.
876 *Science* **342**, 1379-1382 (2013).
877 84. Shi, X. *et al*. Combinatorial GxGxE CRISPR screen identifies SLC25A39 in mitochondrial
878 glutathione transport linking iron homeostasis to OXPHOS. *Nat Commun* **13**, 2483 (2022).
879 85. Du, Y. *et al*. Validation of housekeeping genes as internal controls for studying gene expression
880 during Pacific oyster (Crassostrea gigas) development by quantitative real-time PCR. *Fish*
881 *Shellfish Immunol* **34**, 939-945 (2013).
882 86. Zhang, Q. L. *et al*. Selection of reliable reference genes for normalization of quantitative RT-PCR
883 from different developmental stages and tissues in amphioxus. *Sci Rep* **6**, 37549 (2016).

884

Figures and Figure Legends

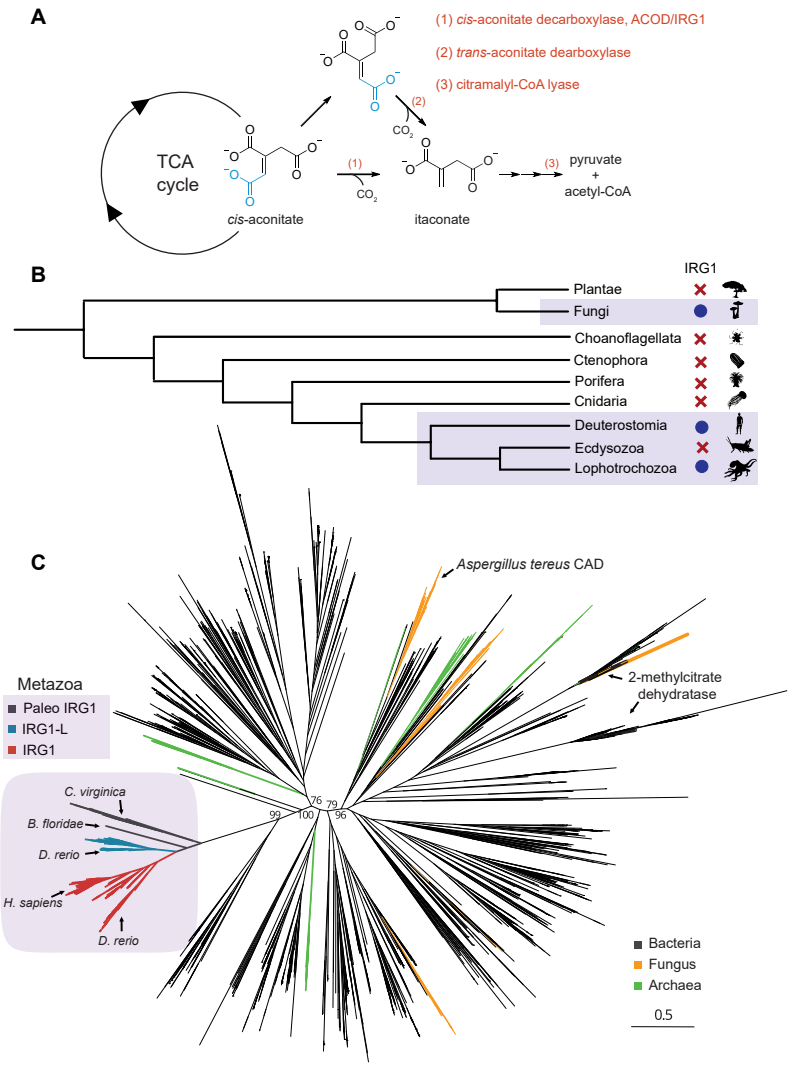


Figure 1. Taxonomic distribution and molecular phylogeny of IRG1, the enzyme producing the innate immune metabolite itaconate. (A) Itaconate biosynthesis and degradation pathways in nature. Two itaconate biosynthesis pathways produce itaconate from the TCA cycle intermediate *cis*-aconitate, through either the *cis*-aconitate decarboxylase (CAD) enzyme IRG1/ACOD1 or the trans-aconitate decarboxylase (TAD) enzyme. Itaconate degradation involves a dedicated citramalyl-CoA lyase CLYBL to generate pyruvate and acetyl-CoA. (B) Taxonomic distribution of IRG1 in eukaryotes. IRG1 is present in Fungi, Lophotrochozoans,

and Deuterostomes but is absent in other metazoans. (C) A phylogenetic tree of IRG1 homologs in nature. Bacterial sequences are colored in black, archaeal in green and fungal in orange. High confidence bootstrapping values are shown for the main branches separating the metazoan and *A.tereus* CAD branches. Branches containing experimentally validated 2-methylcitrate dehydratases are labeled. In the Metazoan branch, sequences were colored based on three lineages, IRG1/ACOD1 (red), IRG1L (IRG1-like, cyan) and Paleo-IRG1 (pIRG1, black). Metazoan proteins that are subsequently studied here are labeled, *C. virginica* (eastern oyster) pIRG1, *B. floridae* (lancelet) pIRG1 and *D.rerio* (zebrafish) IRG1L, in comparison to *D.rerio* (zebrafish) IRG1/ACOD1 and *H.sapiens* (human) IRG1/ACOD1.

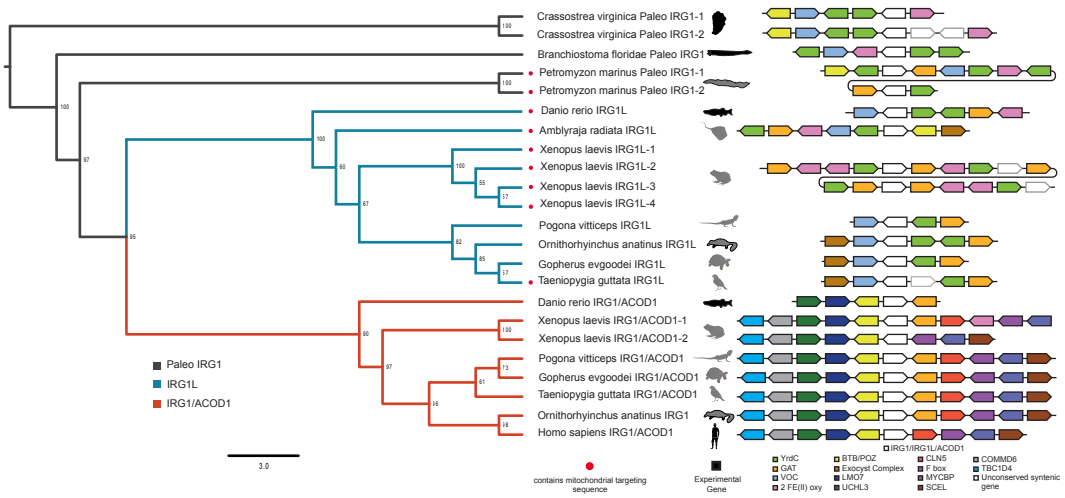


Figure 2. Phylogeny and synteny for representative metazoan IRG1 sequences. The tree branches are color-coded for the three lineages including paleo-IRG1 (p IRG1, black), IRG1-like (IRG1L, cyan) and IRG1 (red). The presence of the predicted Mitochondrial Targeting Sequence (MTS) is highlighted by the red dots. Homologous genes in the genomic vicinity spanning the IRG1, IRG1L and IRG1 are color-coded, for which the homologs present in human were annotated by human gene names, or otherwise by Pfam protein domain name. The two copies of pIRG1 in *C. virginica* are separated by approximately 200 kb, indicated by a dashed line in the syntenic block. Images for experimentally examined species in this report are shown in black, including easter oyster (*C. virginica*), amphioxus (*B. floridae*), and zebrafish (*D. rerio*) IRG1L.

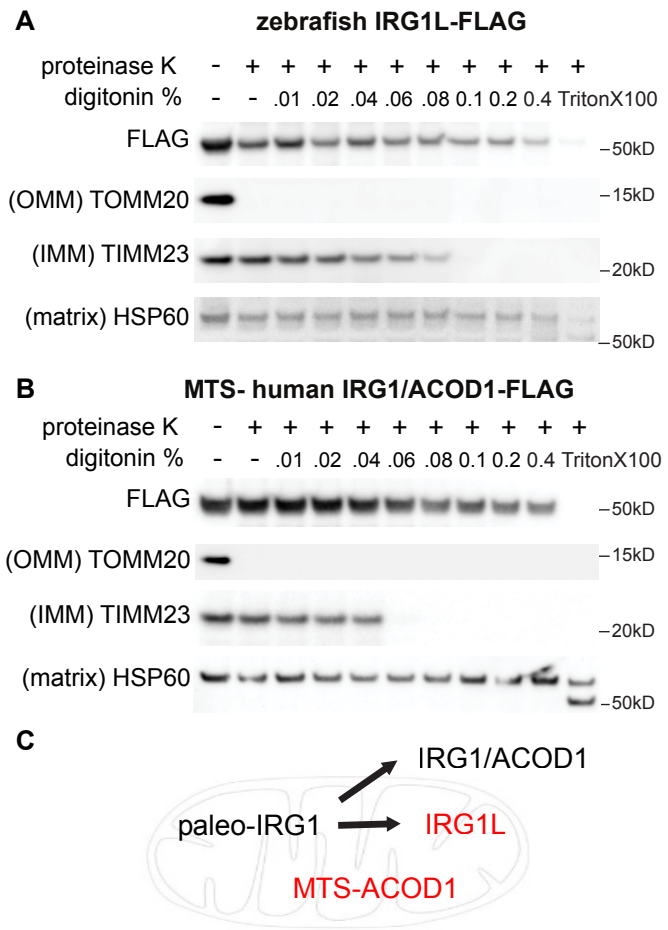


Figure 3. Mitochondrial matrix localization of zebrafish IRG1L. Limited proteolysis of mitochondria from HEK293T cells expressing zebrafish IRG1L (A) and human IRG1 with engineered mitochondrial targeting sequence (MTS) (B). **C**, diagram summarizing the representative submitochondrial localization of paleo-IRG1, IRG1L, IRG1/ACOD1 and engineered MTS-ACOD1.

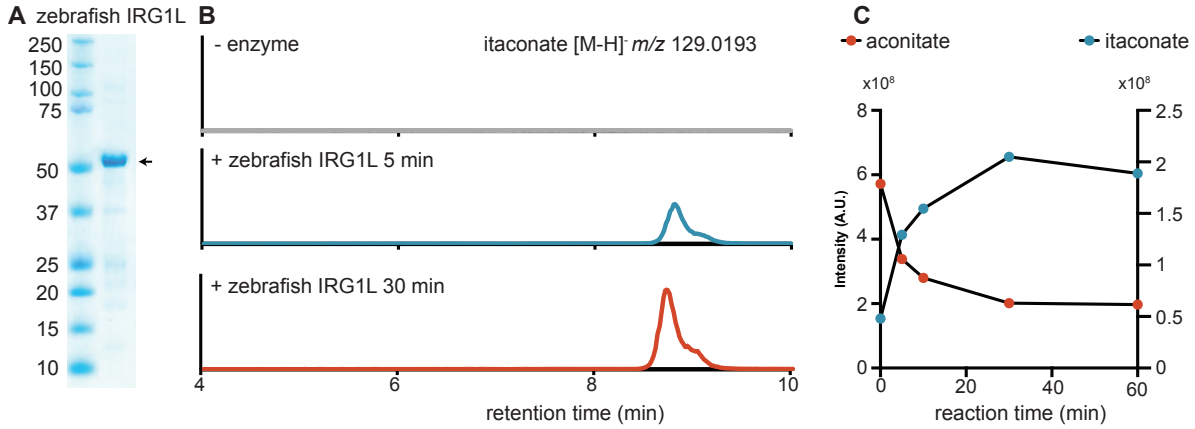


Figure 4. Zebrafish IRG1L is a *cis*-aconitate decarboxylase producing itaconate. (A) Recombinant His-tagged zebrafish IRG1L was expressed from *E. coli* and purified by affinity purification. (B-C) LC-MS analysis confirms itaconate production from *cis*-aconitate (1mM) by zebrafish IRG1L in a time-, enzyme-dependent manner.

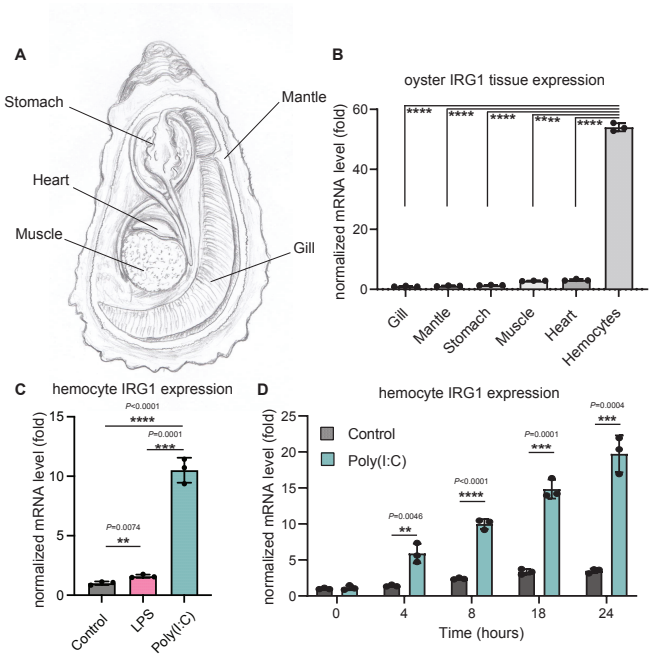


Figure 5. The expression of paleo-IRG1 in eastern oysters is stimulated by Poly(I:C). (A) oyster anatomy. (B) Quantitative PCR showing unstimulated oyster (n = 5) pIRG1 tissue expression ($P_{\text{Hemocytes/other tissue}} < 0.0001$). (C) The pIRG1 expression in the oyster hemocytes is induced Poly(I:C) at 8 hrs shown by qPCR. (D) The time course of pIRG1 expression induction in the oyster hemocytes by Poly(I:C) (10 ug/ml) stimulation. Data are shown as mean ± s.d.

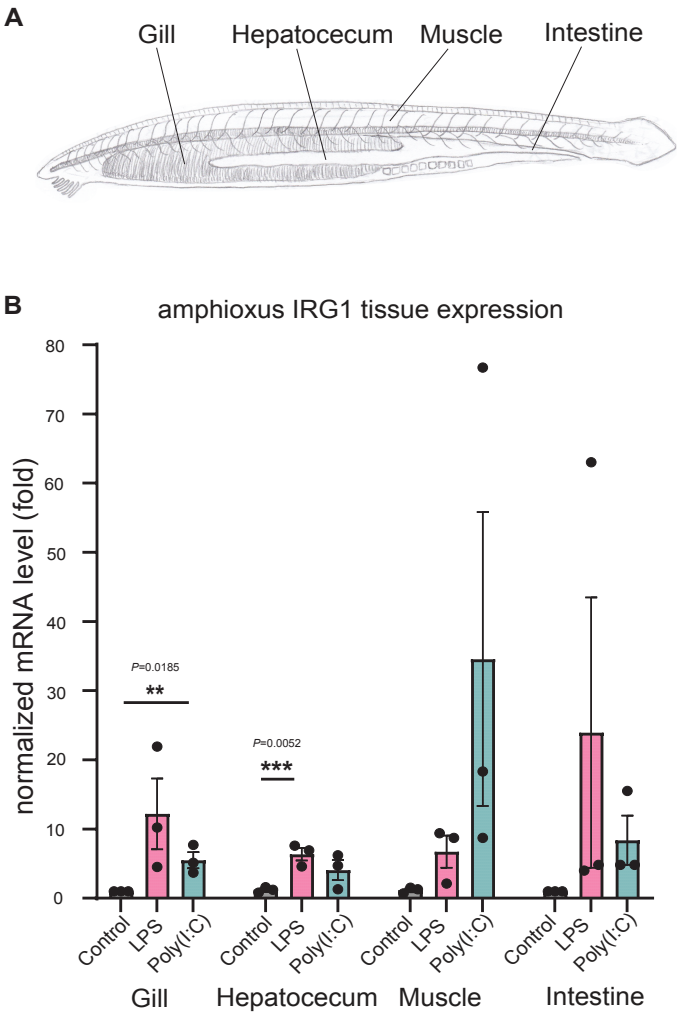
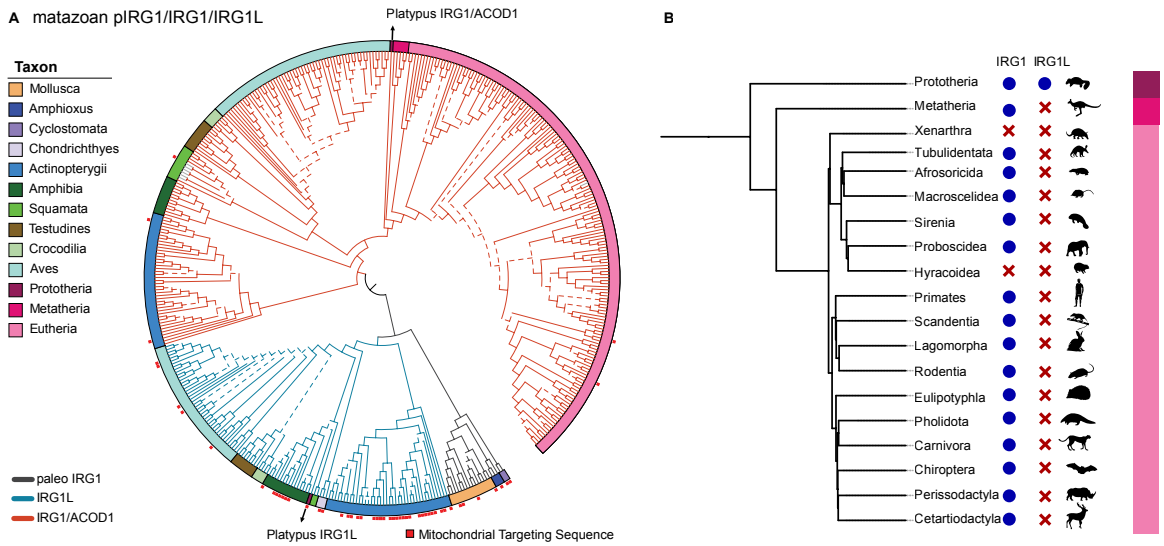
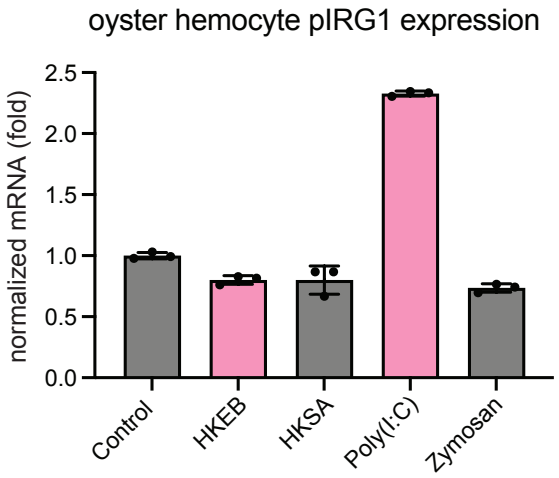


Figure 6. The paleo-IRG1 in amphioxus is implicated in innate immunity. (A) amphioxus anatomy. (B) Quantitative PCR from amphioxus tissues showing that pIRG1 is barely expressed under unstimulated condition (*n* =3), but can be strongly induced by both LPS (50 ug/mL; *n* =3) and Poly (I:C) (1 mg/mL; *n* = 3) upon oral injection for 24 hrs. Each data point represents tissue from one animal shown as mean ± s.d. The large deviance is likely due to the low and variable mRNA level under basal condition.

Supplementary Figures.



Supplementary figure 1. Phylogenetic tree and taxonomic distribution of pIRG1/IRG1L/IRG1 sequences in metazoa. (A) A phylogenetic tree of all metazoan IRG1 sequences. Tree branches were color-coded by three lineages: pIRG1 (black), IRG1L (cyan) and IRG1 (red). Branches with bootstrapping values higher than 70 are highlighted by solid line. Taxa are color-coded on the outer rim of the tree. The sequences with predicted Mitochondrial Targeting Sequence (MTS) were highlighted by red dots. (B) The presence of IRG1 and IRG1L in major mammalian taxa.



Supplementary Figure 2. The pIRG1 is induced in the eastern oyster hemocytes by innate immune stimuli. Quantitative PCR showing the IRG1 mRNA level in hemocytes upon 8 hr stimulation of heat-killed *E. coli* 0111:B4 (HKEB) ($1x10^7$ PFU*)*, heat-killed *S. aureus* (HKSA) ($1x10^7$ PFU), zymosan (100 µg/mL) and Poly (I:C) (10 µg/mL). IRG1 is significantly upregulated upon Poly(I:C) stimulation relative to other conditions.

Supplementary File 1 The sequences used to for phylogenetic inference (.FASTA files)