# Molecular evolution of IRG1 shapes itaconate production in metazoans and alleviates the "double-edged dilemma" of innate immune defense

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 Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

**Abstract**

Itaconate is an innate immune metabolite specifically produced in stimulated immune cells via the decarboxylation of the TCA cycle intermediate *cis*-aconitate. Due to its inhibition of succinate-related metabolic processes, itaconate exhibits antimicrobial properties at the expense of potentially disrupting the hosts' own central energy metabolism, a "double-edged dilemma" of immunometabolism. To understand the evolutionary logic of itaconate biosynthesis, we investigate the evolutionary trajectory of the Irg1 gene, which codes for the itaconate-producing enzyme *cis*-aconitate decarboxylase (CAD). Phylogenetic analysis reveals a putative independent acquisition of metazoan and fungal Irg1 from prokaryotic sources. The metazoan Irg1 underwent duplication in vertebrates and a subsequent loss of one paralog in mammals, a process that removes the mitochondrial targeting sequences (MTS) and relocates CAD outside of the mitochondrial matrix thus preventing a direct inhibition of energy metabolism. Experiments in the most diverged metazoan species that are known to contain Irg1, oysters and amphioxus, reveal that the expression of primitive Irg1 genes is induced by innate immune stimulants in invertebrates and basal chordates, suggesting an already specialized role of itaconate production for innate immune defense in early bilaterians. Our combined *in silico* and experimental analysis of Irg1 highlights that a trend of tightened transcriptional regulation and sequence-level change optimizes itaconate biosynthesis for innate immunity, a mechanism that may be broadly utilized to resolve other types of double-edged dilemmas in immunometabolism.

## Introduction

47 Recent multispecies analysis reveals that enzyme biochemistry dedicated to modern eukaryotic
48 innate immunity has deep evolutionary roots, with an already specialized role in the innate
49 immunity of ancient life: examples include the cGAS-STING pathway in animals that shares
50 ancestry with the prokaryotic cyclic-oligonucleotide in anti-phage defense[1-4], and the eukaryotic
51 peptidoglycan amidase that digests the bacterial cell wall and was horizontally acquired from
52 prokaryotic antibacterial toxins[5]. We rationalize that tracing the evolutionary trajectory of the
53 specialized biochemistry in innate immunity might provide insights on the regulatory
54 mechanisms that shape and optimize the modern innate immune response compatible with the
55 organismal fitness.
56
57 Itaconate, an innate immunometabolite with both antimicrobial and anti-inflammatory roles[7,8] 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)[9], shunting away the central energy metabolic pathway TCA
61 cycle intermediate *cis*-aconitate at a high flux rate (Figure 1a). Itaconate is highly reactive due to
62 the terminal double bond: the alkene group on the C2 position (Figure 1a). This reactivity
63 enables antimicrobial activity via inhibiting isocitrate lyase in the glyoxylate shunt, a key energy
64 metabolic pathway exclusively found in prokaryotes[9-11], and vitamin B12-dependent
65 methylmalonyl-CoA mutase-mediated odd carbon utilization[12,13]. However, in the host, itaconate
66 can inhibit key mitochondrial matrix-localized energy metabolic steps including succinyl-CoA
67 ligase[14], succinate dehydrogenase[15,16] and B12-dependent methylmalonyl-CoA mutase producing
68 succinate[12], in addition to functioning as a Michael acceptor reacting to nucleophilic sulfur group
69 of the cysteine residue of KEAP1 activating Nrf2[17] and that of the antioxidant glutathione[18].

70 Itaconate can be completely degraded by a citramalyl-CoA lyase activity (Figure 1A) that was
71 discovered for human CLYBL gene protecting the vitamin B12 cofactor[12]. The same activity of
72 bacterial citE provided fitness (virulence) to the bacterial pathogen by degrading itaconate to
73 counter the host defense[19], which illustrates an evolutionary arms race.
74
75 Taken together, itaconate's dual inhibition of the host and pathogen central energy metabolic
76 pathway presents a "double-edged" dilemma in immunometabolism: the host benefits from the
77 elimination of pathogens, but also suffers an apparent detrimental effect via the disruption of the
78 organism's own energy metabolism. We refer to this trade-off as the "double-edged" dilemma
79 for the toxic immunometabolic biochemistry in immunity.
80
81 One mechanism to alleviate this dilemma involves a spatiotemporal restricted transcriptional
82 regulation of the Irg1 gene, where its expression is exclusively restricted to macrophages in
83 humans, mice[9], and zebrafish[20]. Within macrophages, Irg1 transcription is strictly kept silent and
84 only upregulated upon innate immune stimulation by bacterial pathogens or LPS[6]. Only a few
85 studies reported the induction of Irg1 in other cell types by other stimuli, including Irg1 paralog
86 in zebrafish that is induced in terminally differentiated keratinocytes upon bacterial infection[21]
87 and 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 distantly diverged species where the mechanism of transcriptional
91 regulation might be less sophisticated.
92

93 To explore the evolutionary logic of utilizing itaconate production for innate immunity, we
94 investigated the taxonomic distribution and the evolutionary trajectory of Irg1 *in silico*. We
95 discovered that fungal Irg1 and metazoan Irg1 are likely independently acquired from
96 prokaryotes. Whole genome duplication early in vertebrate history appears to have produced two
97 paralogous Irg1 copies and subsequent loss events in therian mammal lineages (marsupials and
98 placental mammals) enabled more precise regulation of itaconate biosynthesis by removing the
99 mitochondrial targeting sequence (MTS) in Irg1, thus preventing a direct inhibition of
100 mitochondrial matrix bioenergetics. Our cell- and tissue-based studies in the most distantly
101 diverged metazoan species that contain Irg1, mollusks (eastern oyster, *Crassostrea virginica*)
102 and basal chordates (amphioxus, *Branchiostoma floridae*), identified a transcriptional induction
103 of Irg1 by innate immune inducers, suggesting an ancestrally dedicated innate immune role for
104 Irg1 in ancient metazoans. Together, our molecular phylogeny and experimental studies in these
105 nontraditional model organisms reveal a multilayered evolutionary strategy that tightens the
106 regulations of itaconate biosynthesis. We speculate that similar evolutionary regulatory
107 mechanisms may be broadly utilized to shape toxic innate immune biochemistry and to resolve
108 the "double-edged" dilemma in immune defense.
109

110 **Results**

111 To understand the occurrence of itaconate production in nature, our initial literature survey
112 identified two enzymes that synthesize itaconate: the cis-aconitate decarboxylase (CAD) enzyme
113 in humans, mice, zebrafish[9,20], and the fungus *Aspergillus terreus*[23,24], and *trans*-aconitate
114 decarboxylase (TAD) in *Ustilago maydis*, a plant pathogenic fungus[25] (Figure 1A). Primary
115 sequence analysis and domain searches suggest that all CAD sequences are comprised of the

116 MmgE/PrpD domain fold[26], a protein domain family that also contains the prokaryotic
117 methylcitrate dehydratase in the methylcitrate cycle[27]. The TAD enzymes are comprised of a
118 completely different protein fold that is shared by the adenylosuccinate lyase enzyme ADSL in
119 mammalian purine nucleotide biosynthesis. However, TAD enzymes appear to be exclusively
120 restricted to the fungi kingdom and metazoans lack TAD enzymes. While many fungal studies
121 focus on the bioengineering role of itaconate production for polymer biosynthesis[28], the
122 physiological role of *A. terreus* CAD and *U. maydis* TAD is not completely understood — a
123 potential function in immune defense may hint at convergent evolution to produce itaconate
124 starting from two different protein folds and slightly different substrates[29].
125
126 Because the CAD enzyme exhibits a broad taxonomic distribution spanning fungi and mammals,
127 we next focused on exploring the evolutionary trajectory of CAD across the tree of life.
128 Combined BLAST searches (BLASTP and TBLASTN searches using human Irg1 protein
129 sequence as a query) and a MmgE/PrpD Pfam domain search returned protein sequences
130 spanning all three domains of life: bacteria, archaea and eukaryotes. A closer inspection of the
131 eukaryotic domain suggests a sporadic presence of Irg1 homologs in fungi, deuterostomes, and
132 lophotrochozoans, but CAD is notably absent in their sister clades including ecdysozoans
133 (arthropods, nematodes and their relatives) and plants (Figure 1B). Six putative Drosophila
134 sequences were determined to be misannotated and thus excluded for further analysis (see
135 methods). The MmgE/PrpD domain-containing proteins are present in Basidiomycota and
136 Ascomycota, the fungi taxa including *A. terreus*, but are absent in other basal fungi taxa
137 including Chytridiomycota and Zygomycota. The patchiness of Irg1 occurrence in nature

138 suggests that itaconate biosynthesis is not critical, and thus dispensable for eukaryotic central
139 energy metabolism and organismal fitness.
140
141 Molecular phylogenetic analysis of all prokaryotic and eukaryotic Irg1 homologs highlighted a
142 clear separation of the fungi and animal Irg1 sequences, raising the possibility of two
143 independent acquisition events. We first confirmed that all the Irg1 sequences contain the
144 MmgE/PrpD domain and then manually added eight experimentally validated 2-methylcitrate
145 dehydratases including bacterial sequences from *E. coli*[30], *S. typhimurium*[31], *C. necator*[32], *C.*
146 *glutamicum*[33], *B. subtilis*[34], and two different homologs in *M. tuberculosis*[35-37] as well as a fungal
147 sequence from *Y. lipolytica*[38]. The most striking observation from the phylogenetic analysis
148 (Figure 1C) is that, despite similar enzymatic activities, the validated fungi *A. terreus* CAD and
149 the metazoan Irg1 sequences reside on two distant branches intercalated by prokaryotic
150 sequences with high bootstrapping values (>75), strongly suggesting that metazoan CAD/Irg1
151 and the fungi CAD are independently acquired.
152
153 The absence of the CAD enzyme sequences in many fungal taxa and the close relationship of *A.*
154 *terreus* CAD sequence with the prokaryotic sequences in adjacent branches may indicate a
155 putative direct horizontal gene transfer (HGT) into the *Aspergillus* ancestor. The eight
156 experimentally validated 2-methylcitrate dehydratases are clustered on a separate branch from
157 both *A. terreus* CAD and metazoan Irg1, suggesting these are indeed different enzymes that can
158 be distinguished by sequence homology. Notably, the nearest clade to metazoan Irg1 orthologs
159 contain primarily alphaproteobacterial sequences including one from the Rickettsiales genera,

160 the bacterial taxa considered to be the closest extant organism to eukaryotic mitochondria[39].
161 However, from which prokaryotic clades metazoans acquired Irg1 remains unknown.
162
163 The evolutionary trajectory of metazoan Irg1 is consistent with canonical metazoan phylogeny
164 and is composed of three distinct subbranches: a homolog in mollusks and basal chordates we
165 called paleo-Irg1 (pIrg1, Fig. 2, in gray) and two paralogous vertebrate groups that we
166 designated Irg1-like (Irg1L, Fig. 2, in blue) and Irg1 (Fig. 2, in red). The pIrg1 sequences consist
167 of invertebrate sequences exclusively in mollusks, the basal chordate amphioxus (*Branchiostoma*
168 sp.), and agnathans (lamprey, *Petromyzon marinus*) (Figure 1C). A clear divergence of the Irg1L
169 and Irg1 branches in several gnathostome classes (in Actinopterygii, Sarcopterygii, Amphibia,
170 Aves, and Reptilia) and the conspicuous absence of Irg1L in therian mammals prompted us to
171 further inspect the taxonomic distribution and phylogeny of Irg1L/Irg1 in metazoans.
172
173 We then focused on the metazoan tree and color-coded the sequences in the three subbranches
174 corresponding to pIrg1, Irg1 and Irg1L, by taxonomy (Figure S1A). After analyzing the
175 taxonomic distribution of the Irg1 homologs, we confirmed that all Irg1 and Irg1L sequences
176 belong to jawed vertebrates and that each paralogous branch follows a canonical vertebrate
177 phylogeny. The presence of both Irg1 and Irg1L in the majority of jawed vertebrate taxa,
178 including bony fish, amphibians, reptiles, and birds and the exclusive presence of pIrg1 in
179 mollusks and agnathans strongly support that Irg1 and Irg1L are derived from the pIrg1 through
180 the putative genome duplication event at the base of the vertebrate lineage[40].
181

182 Several taxon-specific local gene duplication events were observed, including two syntenic
183 pIrg1s in oysters (>99.48% nucleotide sequence identity), two syntenic pIrg1s in lamprey
184 (>98.72% nucleotide sequence identity), four syntenic Irg1Ls in *Xenopus laevis* (Figure S1 and
185 2), and two Irg1Ls in salmonid fish likely due to known genome duplication in these taxa[41].
186
187 Cartilaginous fish genomes were found to only contain Irg1L, but lack Irg1 (Figure S1B), a
188 result we further confirmed by a manual BLAST search in species including the skate
189 (*Amblyraja radiata*) and sharks (*Chilscyllium plagiosum, Rhincodon typus, Chiloscyllium*
190 *punctatum, Carcharodon carcharias, Scyliorhinus torazame, Callorhinchus milii and*
191 *Scyliorhinus canicula*).
192
193 The most striking observation in the taxa-annotated metazoan tree is a complete loss of Irg1L in
194 the metatherian and eutherian mammals that otherwise exhibit an expansion of Irg1 lineage (Fig.
195 S1, pink), which contrasts with the presence of both Irg1 and Irg1L in the prototherian platypus
196 (*O. anatinus*; Figure S1, arrows). To further confirm the difference, we manually searched for
197 Irg1 and Irg1L in major mammalian orders[42] by either BlastP or TblastN search using human
198 Irg1 and zebrafish Irg1L sequences (Figure S1B). The absence of both Irg1 and Irg1L in
199 Hyracoidea and Xenarthra (Fig S1B) is consistent with the dispensability of Irg1 for organismal
200 energy metabolism. All metatherians and eutherians contain only Irg1. Within the mammalian
201 clade, Irg1L is exclusive to prototheria, a primordial egg-laying mammal, and completely lost in
202 metatheria and eutheria, the viviparous mammals (Figure S1B), inviting further analysis in the
203 difference between Irg1L and Irg1.
204

205 Primary sequence inspection unexpectedly revealed a N-terminus extended amino acid sequence
206 that is frequently present in both Irg1L and pIrg1 sequences, but much less frequent in Irg1.
207 Because the CAD enzyme directly acts on intermediates of the mitochondrial matrix-localized
208 central energy metabolism TCA cycle, we then explored if these N-terminal extensions might
209 encode the canonical mitochondrial targeting sequence (MTS)[43], the peptide motif directing
210 proteins into the mitochondrial matrix. Indeed, investigating all metazoan sequences using MTS
211 prediction software TargetP revealed high-confidence MTS present in 34.56% pIrg1 and Irg1-L
212 sequences, and only 5 present in the Irg1 sequences (<1.59%; Figure S1 and Figure 2). Loss of
213 the MTS in Irg1 and the loss of mitochondria targeted Irg1L in the modern mammals might
214 confer a fitness advantage by redirecting Irg1 to cytoplasm and alleviating itaconate's direct
215 inhibition of the mitochondrial energy metabolism TCA cycle.
216
217 To further establish the paralogy of Irg1L and Irg1, we performed a manual synteny analysis to
218 probe the genomic context surrounding pIrg1, Irg1-L and Irg1 in the representative species from
219 the major metazoan taxa that include well-studied model organisms (Figure 2). This subset of the
220 tree retains both the branch topography (pIrg1, Irg1-L and Irg1) and the enrichment of MTS in
221 the pIrg1 and Irg1L homologs (Figure 2). We then manually annotated and compared the coding
222 genes in the genomic region harboring pIrg1/Irg1L/Irg1. Although we did not detect obvious
223 functionally related gene clustering[44,45] the synteny surrounding Irg1 appears to have converged
224 to a more highly conserved pattern than that of Irg1L, both in the number of the syntenic genes
225 and their directionality (Figure 2). Four closely examined genes syntenic to pIrg1 appear to
226 contain the same protein domains as the genes syntenic to Irg1L, but not the genes syntenic to
227 Irg1, which we labeled based on their conserved protein domains: VOC, YrdC, 2Fe (ii) oxy and

228 GAT (Figure 2). The exception is the POZ domain protein that is found in the oyster synteny but
229 are fused to BTB domain in the human gene KCTD12 that is syntenic to human Irg1. We
230 therefore conclude that pIrg1 has a more similar synteny to Irg1L than to Irg1.
231
232 To confirm that pIrg1 and Irg1L are indeed cis-aconitate decarboxylase and are thus implicated
233 in innate immunity similar to human Irg1, we performed both *in vitro* enzymology to study the
234 recombinant zebrafish Irg1L protein, as well as *ex vivo* cellular and tissue-level analysis of
235 paleo-Irg1 in eastern oysters (*Crassostrea virginica*) and amphioxus (*Branchiostoma floridae*).
236
237 First, *in vitro* enzymology using the zebrafish Irg1L recombinant protein confirmed the same cis-
238 aconitate decarboxylase activity that is reported for human Irg1 and *A. terreus* ACOD[26]. We
239 purified N-terminally His-tagged recombinant zebrafish Irg1L protein (aa 32- 466) lacking the
240 predicted mitochondrial targeting sequence (Figure 3C). Incubating this recombinant protein
241 with aconitate *in vitro* produced itaconate in an enzyme- and time-dependent manner, detected
242 by LC-MS (Figure 3B and 3C).
243
244 To probe a potentially ancient role of pIrg1 in innate immunity, we analyzed pIrg1 in the most
245 distantly diverged metazoan from humans that contain the Irg1 gene, mollusks, specifically
246 eastern oysters (*C. virginica*). Under resting conditions, the pIrg1 mRNA can be detected from
247 several major oyster tissues including gills, mantle, stomach, muscles, and heart (Figure 4A),
248 however, a strikingly high level of Irg1 was detected in the hemocytes (Figure 4B), the cells in
249 the oyster hemolymph extracted from the adductor muscles[46]. The adult oysters contain a semi-
250 open circulatory system in which the hemolymph (analogous to blood) directly bathes other

251 organs systems[47-49]. The oyster hemocytes exhibit dual functionality in both nutrient acquisition
252 and innate immune defense via phagocytic properties, resembling the macrophages in the starfish
253 larvae observed by Elie Metchnikoff[50,51].
254
255 The transcriptional regulation of Irg1 in mammals and the high pIrg1 level in the oyster
256 hemocytes prompted us to explore pIrg1 in innate immunity. We isolated and cultured the oyster
257 hemocytes and measured pIrg1 mRNA level in response to a panel of innate immune inducers,
258 including heat-killed gram-negative bacterial pathogen *E. coli* (HKEB), heat-killed gram-
259 positive bacterial pathogen *S. aureus* (HKSA), zymosan (mimicking fungal infection), and
260 Poly(I:C) (mimicking viral infection) for 8 hours. We found that the oyster paleo-Irg1 was
261 induced by poly(I:C) stimulation (Figure S4 and Figure 4E), but not LPS (mimicking gram-
262 negative bacterial infection) (Figure 4E), and that induction continues to increase up to 24-hr
263 stimulation (Figure 4D). This transcriptional regulation in oysters is different than in mammals,
264 where the Irg1 induction is triggered by LPS with a peak expression 6-hr post stimulation[9].
265
266 We then investigated pIrg1 in the basal chordate amphioxus (*B. floridae*). Amphioxus, or
267 lancelets, belong to the subphylum cephalochordata, which is thought to have shared a common
268 ancestor with more advanced chordates > 500 mya[52]. This 4-7 cm long, translucent, fish-like
269 animal contains a notochord, but not vertebrae, is considered a basal chordate, and is often
270 studied to provide insights into the origins of vertebrates[53]. Lacking circulatory cells, the cellular
271 immunity in the adult animals is thought to be provided by tissue-resident macrophage-like
272 coelomocytes, which have been reported at cellular resolution in the intestine[54]. To probe
273 amphioxus pIrg1 in innate immunity, the adult animals were stimulated with LPS and Poly(I:C)

274 by oral administration, and then recovered in salt water for 24 hrs prior to tissue dissection,
275 mRNA extraction and qPCR analysis for pIrg1 (Figure 5A). pIrg1 is barely detected in the major
276 tissues including the gills, hepatocecum, muscle and intestine in the control animals (Figure 5B),
277 in which the inconsistent detection in some replicates with very large amplification cycle
278 numbers (>35) preclude an accurate quantification for normalization. In comparison, the pIrg1
279 transcripts are dramatically upregulated and consistently detected in all tissues after both LPS
280 and Poly(I:C) stimulation (Figure 5B). A transcriptional induction of pIrg1 by LPS and Poly(I:C)
281 support the role of pIrg1 in amphioxus innate immunity, suggesting that this specialized,
282 itaconate-mediated biochemistry has already been programmed for innate immunity very early in
283 metazoans.
284
285 **Discussion**
286 Our experimental studies of Irg1 in oysters and amphioxus, the most distantly diverged metazoan
287 species that are known to contain Irg1, revealed itaconate biosynthesis is enriched in the innate
288 immune cells (in oysters) and is induced upon innate immune stimulus (in oysters and
289 amphioxus), suggesting an ancient and yet already specialized role of itaconate in innate
290 immunity at the base of the metazoan lineage. The induction of oyster pIrg1 is consistent with
291 the increased itaconate level reported in oysters upon marine herpesvirus infection[55] and is also
292 supported by the induced expression of pIrg1[56] and increased itaconate[57,58] in another marine
293 bivalve (mussels; *Perna canaliculus*) upon infection with a gram-negative *Vibrio* strain. In these
294 contexts, itaconate might play an antimicrobial function, as itaconate inhibits a pathogenic
295 marine *Vibrio* strain growth *in vitro* and disruption of central carbon metabolism[59].
296

297 Taxonomic distribution and evolutionary trajectory of metazoan Irg1 provide hints into the
298 evolutionary logic that tightens the regulation of the specialized itaconate biochemistry for innate
299 immune killing to alleviate the "double-edged" dilemma of immunometabolism. The frequent
300 loss of Irg1 in metazoan taxa, particularly the Irg1L paralog in mammalian taxa, clearly suggests
301 dispensability for the host's organismal central energy metabolism. More importantly, due to
302 potentially detrimental effects of itaconate in inhibiting host energy metabolism, itaconate
303 production might reduce host fitness and may therefore be subject to negative selection and/or
304 stringent regulatory strategies in evolution, a notion prompted by several of our observations in
305 this study.
306
307 First, we identified that the mitochondrial targeting sequence (MTS) is frequently detected in
308 pIrg1 and Irg1L, but is almost completely lost in Irg1, the only gene copy that is retained in the
309 eutherian mammals. The loss of Irg1L in the therians coupled with the loss of MTS in Irg1
310 would divert the catalysis outside of the mitochondria and prevent direct inhibition of the
311 mitochondrial matrix-localized energy metabolism and thus improve host fitness. Despite
312 lacking MTS, mammalian Irg1 remains associated with the mitochondria[60-62], although the
313 molecular mechanism supporting the association is yet to be discovered.
314
315 Second, despite a conserved role of itaconate in metazoan innate immunity, the regulation of Irg1
316 transcriptional induction in mammals is more stringent than that of the basal metazoans with
317 potentially different signaling pathways. While the mammalian Irg1 is almost exclusively
318 restricted to LPS-activated TLR4 and interferon signaling in macrophages[61,63], the pIrg1s in

319 oysters and in amphioxus appear to be transcriptionally regulated with less stringency, including
320 a basally expressed pIrg1 in oysters' tissues and hemocytes without any stimulation.
321
322 We reported a transcriptional induction of oyster pIrg1 exclusively by polyI:C but not LPS
323 stimulation, as well as a broader, dual induction of amphioxus pIrg1 by both poly I:C and LPS
324 stimulation. Our report on poly I:C-induced oyster pIrg1 expression is supported by the
325 transcriptomics analysis in mussel hemocytes upon screening for a panel of pathogen-associated
326 molecular patterns (PAMPs), where mussel pIrg1 induction is restricted to poly I:C stimulation[64].
327 Lophotrochozoa (including marine bivalves) and amphioxus contain rudimentary TLR
328 sequences[65-68], however, it remains unknown if these rudimentary TLRs function as pattern
329 recognition receptors (PRRs) with the same degree of specificity and downstream signaling as
330 their mammalian counterparts[69,70]. Notably, the first identified Drosophila Toll-receptor does not
331 recognize foreign PAMPs despite the involvement in host resistance against fungal and bacterial
332 infection[71]. We speculate that pIrg1 in oysters and amphioxus might be induced by a non-TLR,
333 membrane-bound PRRs. For instance, zebrafish Irg1 may be regulated by cooperation between
334 glucocorticoid receptors and the JAK/STAT pathway[20]. Future multispecies analysis on the
335 immune signaling pathways and transcriptional factors that induce Irg1 expression in non-
336 traditional model organisms is vital to answering these questions.
337
338 Investigation of three metazoan subbranches, pIrg1, Irg1L and Irg1, support the notion that
339 genome duplication and gene loss provide a powerful evolutionary force to innovate gene
340 expression regulatory mechanisms[72]. Our analysis is consistent with a second whole genome
341 duplication in the vertebrate lineage (2R WGD) that produced the paralogs Irg1L and Irg1, in

342 which Irg1L is subsequently lost in viviparous mammalian taxa. The timing of the 2R WGD is
343 controversial; with reports placing either before[73] or after[40] the gnathostome-agnathan divergence
344 occurring around 465-428 Ma. Our findings that lamprey has two pIrg1s due to local gene
345 duplication without either Irg1 and Irg1L sequence would agree that the 2R WGD occurs after
346 the divergence of agnathans and gnathostomes. The relative similarity of Irg1L and pIrg1
347 synteny and the presence of MTS would support the notion that the genome duplication and loss
348 of Irg1L enable an innovation to tighten the regulation of itaconate biosynthesis exerted by extra-
349 mitochondrial Irg1 that is subsequently retained in eutherian mammals.
350
351 Our exploration of the molecular evolution of itaconate biosynthesis presents another case of
352 horizontal gene transfer in eukaryotes from prokaryotic source. The potential independent
353 acquisition of the metazoan Irg1 and the fungi *A. terreus* ACOD suggests at least one, and
354 perhaps two horizontal gene transfer (HGT) events from bacteria to eukaryotes in the
355 evolutionary history of Irg1. The exclusive presence of ACOD in *Aspergillus* taxa within fungi
356 that are nested within other prokaryotic sequences would strongly suggest a HGT in *Aspergillus*.
357 Our studies of the metazoan Irg1 can be explained by two scenarios: first, the metazoan Irg1 is
358 derived from the last common ancestor of eukaryotes and subsequently lost several times in
359 major eukaryotic taxa including plants, most fungi, and ecdysozoa; or two, the metazoan
360 ancestor's Irg1 is horizontally transferred directly from bacteria. The observation that the closest
361 branch next to metazoan Irg1 (Figure 1C) contains mostly sequences from alphaproteobacteria
362 might support the former. However, the frequent local gene duplication in pIrg1 (e.g., in oysters
363 and in sea lampreys) and Irg1L (in xenopus) suggests pIrg1/Irg1L genomic region might be
364 highly mobile, indicative of a more recent HGT event — an analysis of mobile gene elements

365 near pIrg1 would help resolve this question. Notably, a previous systems-level analysis of
366 eukaryotic orthologous groups (KOGs) also highlighted CAD sequences as the top eukaryotic
367 sequences with unexpected phyletic patterns, supporting putative independent HGT events[74]. On
368 the other hand, the study on itaconate production in microorganisms, in addition to the
369 bioengineering application[75], might be insightful. For instance, if the bacterial sequences near the
370 metazoan Irg1 or the *A. terreus* CAD also producing itaconate for bacterial immunity, this would
371 indicate a universal role of itaconate in innate immunity.
372

373 **Methods**

374 **Phylogenetic analysis of the Irg1 Orthologs**

375 Human Irg1 was queried against the NCBI database using BLASTp and tBLASTn[76] manually
376 across the tree of life and were subsampled by taxonomic order during May-July 2020.
377 Sequences from metazoan taxa were restricted to Expectation values below 1e-10. To maximize
378 sequence diversity in more distantly related sequences from fungi, archaea, and prokaryotes
379 inclusion criteria was relaxed to 0.001. In total, 2,297 sequences were identified manually.
380 Duplicate sequences were identified and removed. A full phylogeny of all sequences was
381 constructed using MAFFT (v7.471)[77]. IQTree (v2.1.2)[78] and ModelFinder[79] were used to find the
382 best fit substitution model and branch support values were obtained by ultrafast bootstrapping[80].
383 Figtree (v1.4.4) was used to visualize the resulting tree. A similar method was used to construct
384 phylogenetic trees for the metazoan sequences (supplementary Fig. 1a) and model organism
385 sequences (Fig. 2c) and iTOL (v6.5.2)[81] was used to visualize the tree.
386

387 **Analysis of dubious sequences**

388 Initial sequence surveys revealed only nine orthologs in Arthropoda, all in *Drosophila* genera,
389 three orthologs in Plantae: one in *Ziziphus* and two in *Quercus* generas. Reciprocal BLASTp and
390 phylogenetic analysis indicated likely sequence contamination by *Acetobacter aceti* in
391 Drosophila sequences, a black yeast-like fungal pathogen in the *Quercus* sequence, and *Belnapia*
392 contamination in the *Ziziphus* sequences.
393
394 **Identification of mitochondrial targeting sequences**
395 All the metazoan Irg1 sequences were analyzed using Target-P (v2.0)[82] to predict the presence of
396 N-terminal mitochondrial targeting sequences (MTS). Default settings were used, and a
397 likelihood >0.5 was considered a predicted MTS.
398
399 **Synteny analysis**
400 Representative genes from the model organisms from each major metazoan taxa possessing an
401 Irg1 ortholog were manually investigated for syntenic genes using NCBI Genome Data
402 Viewer[83]. The five nearest annotated genes both upstream and downstream were analyzed using
403 Pfam[84] and NCBI Domain Architecture Retrieval Tool[85] to predict domain architecture. In
404 instances where local duplication is present, more than the five nearest genes were recorded as
405 there were multiple nearby Irg1 homologs (e.g., *Xenopus laevis* in Fig. 2).
406
407 **Zebrafish Irg1-like protein production**
408 A N-terminal His-tagged *Danio rerio* Irg1-like sequence lacking the N-terminus predicted
409 mitochondrial targeting sequence (Δ31aa) was codon-optimized for bacterial expression, custom
410 synthesized and subcloned into pET28 vector (Genscript). The plasmid was transformed in BL21

411 Star™ (DE3) Chemically Competent E. coli (Invitrogen c6010-03), grown at 225 rpm at 37 °C
412 to OD600 0.6. Cells were then induced with 0.1 mM β-D-1-thiogalactopyranoside (IPTG) for 6
413 hours at 18 °C. Cells were lysed by sonication (Qsonica Q500 with microtip) in 200 mL lysis
414 buffer (20mM HEPES pH 7.4, 500 mM NaCl, 10% glycerol v/v, 1 mM TCEP (pH 7.4), 100
415 U/mL DNase, 0.1 mg/mL Lysozyme, and cOmplete mini protease inhibitor cocktail) using 2 sec
416 on and 2 sec off pulses at 25% power for 2 minutes. The recombinant Irg1-like protein was
417 enriched using cobalt TALON Metal Affinity Resin (Takara Bio #635501) with gentle shaking
418 for 1 hour at 4 °C and washed twice by centrifugation for 3 minutes at 3500 rpm at 4 °C and
419 resuspension in the washing buffer (20 mM Tris pH 8.0, 150 mM NaCl, 20 mM Imidazole, 1
420 mM TCEP). Protein was eluted from the affinity resin in the elution buffer (20 mM Tris pH 8.0,
421 150 mM NaCl, 500 mM Imidazole, 1 mM TCEP, 10% glycerol). The recombinant protein was
422 then concentrated using centrifugal filter units (Amicon Ultra-4, Millipore) and dialyzed using
423 Slide-A-Lyzer dialysis cassette into storage buffer (20 mM Tris pH 8.0, 150 mM NaCl, 20%
424 glycerol, 1 mM TCEP) overnight.
425
426 **Enzymology**
427 The enzyme assay for the zebrafish Irg1-like was performed by incubating 7 µg recombinant
428 protein with 1 mM cis-aconitate (Sigma A3412-1G) in 300 µl reaction buffer (10 mM HEPES,
429 150 mM NaCl, 0.1 mM TCEP, 10% glycerol, pH 6.5) at 37 °C. Aliquots of 50 µl were taken at
430 0, 10, 30, and 60 minutes and were immediately quenched with 2 % formic acid, vortexed, and
431 centrifuged at 14,000 rpm for 20 minutes at 4 °C. The supernatant was then transferred to a new
432 centrifuge tube and mixed 1:9 with a 3:1 acetonitrile:methanol solvent for LC-MS analysis

433 following the established polar metabolite profiling protocol in the lab[86]. Itaconate production
434 was confirmed using chemical standard itaconic acid (Sigma Aldrich #I29204).
435
436 **Mollusk Irg1 expression**
437 Eastern oysters (*Crassostrea virginica*, Woods Hole Marine Biological Laboratory) were
438 acclimated in filtered sea water overnight after arrival. To determine unstimulated basal pIrg1
439 expression, five oysters were dissected on ice and 50 mg wet tissue mass from the gill, mantle,
440 stomach, adductor muscle, and heart were collected. Tissue was washed in PBS, then
441 homogenized in RL+BME buffer (Qiagen). mRNA extraction was performed per manufacturer's
442 instruction using the RNeasy mini kit (Qiagen). The mRNA concentration was determined using
443 a Nanodrop and cDNA libraries were prepared using SuperScriptIII First-Strand Synthesis Super
444 Mix for qRT-PCR (Thermo Fisher Scientific).
445
446 **Mollusk genes quantitative PCR**
447 Candidate control gene EF1a was identified from the literature[87]. Primers for EF1a and pIrg1
448 (Table 1) were designed using primer-BLAST on the NCBI server. Primers were ordered from
449 Integrated DNA technologies and pilot qPCR and Sanger sequencing experiments were
450 conducted to validate the primers for pIrg1 and EF1a. qPCR was performed using PowerUp
451 SYBR Green Master Mix (Thermo Fisher A25742) in a 96-well format with Thermo Fisher
452 QuantStudio 6Flex instrument.
453
454 **Mollusk Hemocyte Assays**

455 For all hemocyte assays, live eastern oysters were notched and hemolymph extracted by syringe
456 from the adductor muscles following published protocol[46]. Hemolymph from all oysters was
457 pooled on ice, diluted 1:2 with decontamination washing buffer (PBS, 100 µg/mL penicillin-
458 streptomycin), pipetted into Corning Falcon 60 mm tissue culture plates (Thermo Fisher) and
459 were incubated in the dark at room temperature for 30 min. Cells were then inspected under the
460 microscope to ensure adhesion, and decontamination buffer was aspirated and replaced with 5
461 mL culture media (1:1 OptiMEM media: filtered sea water, 50 µg/mL penicillin-streptomycin).
462 The cultured hemocytes were treated with zymosan (100 µg/mL, Invivogen tlrl-zyd), heat-killed
463 *E. coli* ($1x10^7$ PFU, Invivogen tlrl-hkeb2), heat-killed *S. aureus* ($1x10^7$ PFU, Invivogen tlrl-
464 hksa), and Poly(I:C) (10 µg/mL final concentration, Invivogen tlrl-pic), and filtered sea water
465 (control) for 8 hrs in the dark at room temperature approximately 20 °C (Supplementary Figure
466 2), or Poly(I:C) and LPS E. coli 0111:B4 (1 µg/mL final concentration, Sigma LPS25) for
467 indicated time. After treatments, cells were directly lysed and analyzed as described above.
468
469 **Amphioxus Irg1 expression**
470 Amphioxus (*Branchiostoma floridae*, Gulf Specimen Marine Lab) were kept in filtered sea water
471 in dark conditions at room temperature. Three amphioxus each condition were treated by oral
472 injection with LPS E. coli 0111:B4 (50 µg/mL final concentration, Sigma LPS25), Poly(I:C)(1
473 mg/mL, Invivogen tlrl-pic), or control (filtered seawater). Specifically, amphioxus were cold-
474 immobilized and a 25 gauge needle was used to carefully administer 100 µl stimulus solution
475 through the oral hood. Individuals were immediately placed back into their containers at room
476 temperature and recovered overnight. Individuals were netted, cold immobilized and dissected

477 24 hr post treatment. Gill, muscle, hepatocecum and intestine tissue were collected and washed
478 in PBS before tissue lysis and mRNA expression analysis.
479 Candidate reference genes were identified from literature[88]. Primers for pIrg1 and candidate
480 reference genes S20 and EF1a (Table 1) were designed using NCBI primer-BLAST. Primers
481 were ordered from Integrated DNA technologies, and pilot qPCR and Sanger sequencing
482 experiment were conducted to validate the primers.
483
484
485 **Author Contributions**
486 H.S. and G.W. conceived the initial idea. R.S., A.S., and M.D. performed bioinformatics
487 analysis; H.S., R.S., and F.M. performed recombinant protein production and enzymatic assays;
488 R.S. and H.S. performed dissections and qRT-PCR assays. H.S. and R.S. wrote the initial
489 manuscript with input from G.W. All authors reviewed and approved the manuscript.
490
491 **Acknowledgements**
492 We thank Yale School of Medicine, Yale West Campus, and Yale Systems Biology Institute for
493 the instrumentation support; X. Shi for technical support on the LC-MS; J. Liu for technical
494 support; A. Chavan and L. Wu on assays in amphioxus and oyster; K. Byrne for her insight in
495 sequence analysis; and members of the Shen lab, D. Stadtmauer in the Wagner lab, M. Murphy,
496 J. MacMicking, A. Goodman and J. Galán for fruitful discussions and feedback. The work is
497 supported by Yale School of Medicine startup fund and Yale Systems Biology Institute pilot
498 award. H.S. is also supported by Klingenstein-Simons fellowship Awards in Neuroscience and
499 1907 Foundation Trailblazer Award.

500 **References:**


501 1. Wu, J. & Chen, Z. J. Innate immune sensing and signaling of cytosolic nucleic acids.
502 *Annu Rev Immunol* 32, 461-488, doi:10.1146/annurev-immunol-032713-120156 (2014).
503 2. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-
504 oligonucleotide-based anti-phage signalling systems. *Nat Microbiol* 5, 1608-1615,
505 doi:10.1038/s41564-020-0777-y (2020).
506 3. Duncan-Lowey, B. & Kranzusch, P. J. CBASS phage defense and evolution of antiviral
507 nucleotide signaling. *Curr Opin Immunol* 74, 156-163, doi:10.1016/j.coi.2022.01.002
508 (2022).
509 4. Wu, X. *et al.* Molecular evolutionary and structural analysis of the cytosolic DNA sensor
510 cGAS and STING. *Nucleic Acids Res* 42, 8243-8257, doi:10.1093/nar/gku569 (2014).
511 5. Chou, S. *et al.* Transferred interbacterial antagonism genes augment eukaryotic innate
512 immune function. *Nature* 518, 98-101, doi:10.1038/nature13965 (2015).
513 6. Strelko, C. L. *et al.* Itaconic acid is a mammalian metabolite induced during macrophage
514 activation. *J Am Chem Soc* 133, 16386-16389, doi:10.1021/ja2070889 (2011).
515 7. O'Neill, L. A. J. & Artyomov, M. N. Itaconate: the poster child of metabolic
516 reprogramming in macrophage function. *Nat Rev Immunol* 19, 273-281,
517 doi:10.1038/s41577-019-0128-5 (2019).
518 8. Hooftman, A. & O'Neill, L. A. J. The Immunomodulatory Potential of the Metabolite
519 Itaconate. *Trends Immunol* 40, 687-698, doi:10.1016/j.it.2019.05.007 (2019).
520 9. Michelucci, A. *et al.* Immune-responsive gene 1 protein links metabolism to immunity by
521 catalyzing itaconic acid production. *Proc Natl Acad Sci U S A* 110, 7820-7825,
522 doi:10.1073/pnas.1218599110 (2013).

523 10. McFadden, B. A. & Purohit, S. Itaconate, an isocitrate lyase-directed inhibitor in
524 Pseudomonas indigofera. *J Bacteriol* 131, 136-144, doi:10.1128/jb.131.1.136-144.1977
525 (1977).
526 11. Patel, T. R. & McFadden, B. A. Caenorhabditis elegans and Ascaris suum: inhibition of
527 isocitrate lyase by itaconate. *Exp Parasitol* 44, 262-268, doi:10.1016/0014-
528 4894(78)90107-8 (1978).
529 12. Shen, H. *et al.* The Human Knockout Gene CLYBL Connects Itaconate to Vitamin B12.
530 *Cell* 171, 771-782 e711, doi:10.1016/j.cell.2017.09.051 (2017).
531 13. Ruetz, M. *et al.* Itaconyl-CoA forms a stable biradical in methylmalonyl-CoA mutase and
532 derails its activity and repair. *Science* 366, 589-593, doi:10.1126/science.aay0934 (2019).
533 14. Nemeth, B. *et al.* Abolition of mitochondrial substrate-level phosphorylation by itaconic
534 acid produced by LPS-induced Irg1 expression in cells of murine macrophage lineage.
535 *FASEB J* 30, 286-300, doi:10.1096/fj.15-279398 (2016).
536 15. Cordes, T. *et al.* Immunoresponsive Gene 1 and Itaconate Inhibit Succinate
537 Dehydrogenase to Modulate Intracellular Succinate Levels. *J Biol Chem* 291, 14274-
538 14284, doi:10.1074/jbc.M115.685792 (2016).
539 16. Lampropoulou, V. *et al.* Itaconate Links Inhibition of Succinate Dehydrogenase with
540 Macrophage Metabolic Remodeling and Regulation of Inflammation. *Cell Metab* 24,
541 158-166, doi:10.1016/j.cmet.2016.06.004 (2016).
542 17. Mills, E. L. *et al.* Itaconate is an anti-inflammatory metabolite that activates Nrf2 via
543 alkylation of KEAP1. *Nature* 556, 113-117, doi:10.1038/nature25986 (2018).

544 18. Bambouskova, M. *et al.* Electrophilic properties of itaconate and derivatives regulate the
545 IkappaBzeta-ATF3 inflammatory axis. *Nature* 556, 501-504, doi:10.1038/s41586-018-
546 0052-z (2018).
547 19. Sasikaran, J., Ziemski, M., Zadora, P. K., Fleig, A. & Berg, I. A. Bacterial itaconate
548 degradation promotes pathogenicity. *Nat Chem Biol* 10, 371-377,
549 doi:10.1038/nchembio.1482 (2014).
550 20. Hall, C. J. *et al.* Immunoresponsive gene 1 augments bactericidal activity of macrophage-
551 lineage cells by regulating beta-oxidation-dependent mitochondrial ROS production. *Cell*
552 *Metab* 18, 265-278, doi:10.1016/j.cmet.2013.06.018 (2013).
553 21. Hall, C. J. *et al.* Epidermal cells help coordinate leukocyte migration during inflammation
554 through fatty acid-fuelled matrix metalloproteinase production. *Nat Commun* 5, 3880,
555 doi:10.1038/ncomms4880 (2014).
556 22. Daniels, B. P. *et al.* The Nucleotide Sensor ZBP1 and Kinase RIPK3 Induce the Enzyme
557 IRG1 to Promote an Antiviral Metabolic State in Neurons. *Immunity* 50, 64-76 e64,
558 doi:10.1016/j.immuni.2018.11.017 (2019).
559 23. Kanamasa, S., Dwiarti, L., Okabe, M. & Park, E. Y. Cloning and functional
560 characterization of the cis-aconitic acid decarboxylase (CAD) gene from Aspergillus
561 terreus. *Appl Microbiol Biotechnol* 80, 223-229, doi:10.1007/s00253-008-1523-1 (2008).
562 24. Li, A. *et al.* A clone-based transcriptomics approach for the identification of genes
563 relevant for itaconic acid production in Aspergillus. *Fungal Genet Biol* 48, 602-611,
564 doi:10.1016/j.fgb.2011.01.013 (2011).
565 25. Geiser, E. *et al.* Ustilago maydis produces itaconic acid via the unusual intermediate
566 trans-aconitate. *Microb Biotechnol* 9, 116-126, doi:10.1111/1751-7915.12329 (2016).

567 26. Chen, F. *et al.* Crystal structure of cis-aconitate decarboxylase reveals the impact of
568 naturally occurring human mutations on itaconate synthesis. *Proc Natl Acad Sci U S A*
569 116, 20644-20654, doi:10.1073/pnas.1908770116 (2019).
570 27. Dolan, S. K. & Welch, M. The Glyoxylate Shunt, 60 Years On. *Annu Rev Microbiol* 72,
571 309-330, doi:10.1146/annurev-micro-090817-062257 (2018).
572 28. Wierckx, N. *et al.* Metabolic specialization in itaconic acid production: a tale of two
573 fungi. *Curr Opin Biotechnol* 62, 153-159, doi:10.1016/j.copbio.2019.09.014 (2020).
574 29. Galperin, M. Y. & Koonin, E. V. Divergence and convergence in enzyme evolution. *J*
575 *Biol Chem* 287, 21-28, doi:10.1074/jbc.R111.241976 (2012).
576 30. Brock, M., Maerker, C., Schutz, A., Volker, U. & Buckel, W. Oxidation of propionate to
577 pyruvate in Escherichia coli. Involvement of methylcitrate dehydratase and aconitase.
578 *Eur J Biochem* 269, 6184-6194, doi:10.1046/j.1432-1033.2002.03336.x (2002).
579 31. Horswill, A. R. & Escalante-Semerena, J. C. Salmonella typhimurium LT2 catabolizes
580 propionate via the 2-methylcitric acid cycle. *J Bacteriol* 181, 5615-5623,
581 doi:10.1128/JB.181.18.5615-5623.1999 (1999).
582 32. Bramer, C. O. & Steinbuchel, A. The methylcitric acid pathway in Ralstonia eutropha:
583 new genes identified involved in propionate metabolism. *Microbiology (Reading)* 147,
584 2203-2214, doi:10.1099/00221287-147-8-2203 (2001).
585 33. Claes, W. A., Puhler, A. & Kalinowski, J. Identification of two prpDBC gene clusters in
586 Corynebacterium glutamicum and their involvement in propionate degradation via the 2-
587 methylcitrate cycle. *J Bacteriol* 184, 2728-2739, doi:10.1128/JB.184.10.2728-2739.2002
588 (2002).

589 34. Reddick, J. J. *et al.* First Biochemical Characterization of a Methylcitric Acid Cycle from
590 Bacillus subtilis Strain 168. *Biochemistry* 56, 5698-5711,
591 doi:10.1021/acs.biochem.7b00778 (2017).
592 35. Savvi, S. *et al.* Functional characterization of a vitamin B12-dependent methylmalonyl
593 pathway in Mycobacterium tuberculosis: implications for propionate metabolism during
594 growth on fatty acids. *J Bacteriol* 190, 3886-3895, doi:10.1128/JB.01767-07 (2008).
595 36. Griffin, J. E. *et al.* Cholesterol catabolism by Mycobacterium tuberculosis requires
596 transcriptional and metabolic adaptations. *Chem Biol* 19, 218-227,
597 doi:10.1016/j.chembiol.2011.12.016 (2012).
598 37. Munoz-Elias, E. J., Upton, A. M., Cherian, J. & McKinney, J. D. Role of the
599 methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and
600 virulence. *Mol Microbiol* 60, 1109-1122, doi:10.1111/j.1365-2958.2006.05155.x (2006).
601 38. Uchiyama, H., Ando, M., Toyonaka, Y. & Tabuchi, T. Subcellular localization of the
602 methylcitric-acid-cycle enzymes in propionate metabolism of Yarrowia lipolytica. *Eur J*
603 *Biochem* 125, 523-527, doi:10.1111/j.1432-1033.1982.tb06713.x (1982).
604 39. Andersson, S. G. et al. The genome sequence of Rickettsia prowazekii and the origin of
605 mitochondria. Nature 396, 133-140, doi:10.1038/24094 (1998).
606 40. Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate
607 evolution. *Nat Ecol Evol* 4, 820-830, doi:10.1038/s41559-020-1156-z (2020).
608 41. Berthelot C, et al. The rainbow trout genome provides novel insights into evolution after
609 whole-genome duplication in vertebrates. Nat Commun 5, 3657 (2014).

610 42. Prasad, A. B., Allard, M. W., Program, N. C. S. & Green, E. D. Confirming the
611 phylogeny of mammals by use of large comparative sequence data sets. *Mol Biol Evol* 25,
612 1795-1808, doi:10.1093/molbev/msn104 (2008).
613 43. Omura, T. Mitochondria-targeting sequence, a multi-role sorting sequence recognized at
614 all steps of protein import into mitochondria. *J Biochem* 123, 1010-1016,
615 doi:10.1093/oxfordjournals.jbchem.a022036 (1998).
616 44. Osbourn, A. E. & Field, B. Operons. *Cell Mol Life Sci* 66, 3755-3775,
617 doi:10.1007/s00018-009-0114-3 (2009).
618 45. Nutzmann, H. W., Scazzocchio, C. & Osbourn, A. Metabolic Gene Clusters in
619 Eukaryotes. *Annu Rev Genet* 52, 159-183, doi:10.1146/annurev-genet-120417-031237
620 (2018).
621 46. Potts, R. W. A., Gutierrez, A. P., Cortes-Araya, Y., Houston, R. D. & Bean, T. P.
622 Developments in marine invertebrate primary culture reveal novel cell morphologies in
623 the model bivalve Crassostrea gigas. *PeerJ* 8, e9180, doi:10.7717/peerj.9180 (2020).
624 47. Canesi, L., Gallo, G., Gavioli, M. & Pruzzo, C. Bacteria-hemocyte interactions and
625 phagocytosis in marine bivalves. *Microsc Res Tech* 57, 469-476, doi:10.1002/jemt.10100
626 (2002).
627 48. Hine, P. M. The inter-relationships of bivalve haemocytes. *Fish Shellfish Immun* 9, 367-
628 385, doi:DOI 10.1006/fsim.1998.0205 (1999).
629 49. Ford, S. E., Ashtonalcox, K. A. & Kanaley, S. A. Comparative Cytometric and
630 Microscopic Analyses of Oyster Hemocytes. *J Invertebr Pathol* 64, 114-122, doi:DOI
631 10.1006/jipa.1994.1079 (1994).
632 50. Metchnikoff, E. & Binnie, F. G. *Immunity in infective diseases*. (University Press, 1905).

633 51. Tauber, A. I. Metchnikoff and the phagocytosis theory. *Nat Rev Mol Cell Biol* 4, 897-
634 901, doi:10.1038/nrm1244 (2003).
635 52. Chen, J. Y., Huang, D. Y. & Li, C. W. An early Cambrian craniate-like chordate. *Nature*
636 402, 518-522, doi:Doi 10.1038/990080 (1999).
637 53. Huang, S. *et al.* Genomic analysis of the immune gene repertoire of amphioxus reveals
638 extraordinary innate complexity and diversity. *Genome Res* 18, 1112-1126,
639 doi:10.1101/gr.069674.107 (2008).
640 54. Rhodes, C. P., Ratcliffe, N. A. & Rowley, A. F. Presence of coelomocytes in the
641 primitive chordate amphioxus (Branchiostoma lanceolatum). *Science* 217, 263-265,
642 doi:10.1126/science.7089565 (1982).
643 55. Young, T. *et al.* Differential expression of novel metabolic and immunological
644 biomarkers in oysters challenged with a virulent strain of OsHV-1. *Dev Comp Immunol*
645 73, 229-245, doi:10.1016/j.dci.2017.03.025 (2017).
646 56. Sendra, M., Saco, A., Rey-Campos, M., Novoa, B. & Figueras, A. Immune-responsive
647 gene 1 (IRG1) and dimethyl itaconate are involved in the mussel immune response. *Fish*
648 *Shellfish Immunol* 106, 645-655, doi:10.1016/j.fsi.2020.07.034 (2020).
649 57. Nguyen, T. V., Alfaro, A. C., Young, T., Ravi, S. & Merien, F. Metabolomics Study of
650 Immune Responses of New Zealand Greenshell Mussels (Perna canaliculus) Infected
651 with Pathogenic Vibrio sp. *Mar Biotechnol (NY)* 20, 396-409, doi:10.1007/s10126-018-
652 9804-x (2018).
653 58. Nguyen, T. V., Alfaro, A. C., Merien, F., Young, T. & Grandiosa, R. Metabolic and
654 immunological responses of male and female new Zealand Greenshell (TM) mussels

655 (Perna canaliculus) infected with Vibrio sp. J Invertebr Pathol 157, 80-89,
656 doi:10.1016/j.jip.2018.08.008 (2018).
657 59. Nguyen, T. V. *et al.* Itaconic acid inhibits growth of a pathogenic marine Vibrio strain: A
658 metabolomics approach. *Sci Rep* 9, 5937, doi:10.1038/s41598-019-42315-6 (2019).
659 60. Chen, M. *et al.* Itaconate is an effector of a Rab GTPase cell-autonomous host defense
660 pathway against Salmonella. *Science* 369, 450-455, doi:10.1126/science.aaz1333 (2020).
661 61. Tallam, A. *et al.* Gene Regulatory Network Inference of Immunoresponsive Gene 1
662 (IRG1) Identifies Interferon Regulatory Factor 1 (IRF1) as Its Transcriptional Regulator
663 in Mammalian Macrophages. *PLoS One* 11, e0149050,
664 doi:10.1371/journal.pone.0149050 (2016).
665 62. Degrandi, D., Hoffmann, R., Beuter-Gunia, C. & Pfeffer, K. The proinflammatory
666 cytokine-induced IRG1 protein associates with mitochondria. *J Interferon Cytokine Res*
667 29, 55-67, doi:10.1089/jir.2008.0013 (2009).
668 63. Naujoks, J. *et al.* IFNs Modify the Proteome of Legionella-Containing Vacuoles and
669 Restrict Infection Via IRG1-Derived Itaconic Acid. *PLoS Pathog* 12, e1005408,
670 doi:10.1371/journal.ppat.1005408 (2016).
671 64. Moreira, R. *et al.* Stimulation of Mytilus galloprovincialis Hemocytes With Different
672 Immune Challenges Induces Differential Transcriptomic, miRNomic, and Functional
673 Responses. *Front Immunol* 11, 606102, doi:10.3389/fimmu.2020.606102 (2020).
674 65. Zhang, L., Li, L. & Zhang, G. A Crassostrea gigas Toll-like receptor and comparative
675 analysis of TLR pathway in invertebrates. *Fish Shellfish Immunol* 30, 653-660,
676 doi:10.1016/j.fsi.2010.12.023 (2011).

677 66. Gerdol, M. & Venier, P. An updated molecular basis for mussel immunity. *Fish Shellfish*
678 *Immunol* 46, 17-38, doi:10.1016/j.fsi.2015.02.013 (2015).
679 67. Huang, S. *et al.* The evolution and regulation of the mucosal immune complexity in the
680 basal chordate amphioxus. *J Immunol* 186, 2042-2055, doi:10.4049/jimmunol.1001824
681 (2011).
682 68. Sasaki, N., Ogasawara, M., Sekiguchi, T., Kusumoto, S. & Satake, H. Toll-like receptors
683 of the ascidian Ciona intestinalis: prototypes with hybrid functionalities of vertebrate
684 Toll-like receptors. *J Biol Chem* 284, 27336-27343, doi:10.1074/jbc.M109.032433
685 (2009).
686 69. Beutler, B. & Rietschel, E. T. Innate immune sensing and its roots: the story of
687 endotoxin. *Nat Rev Immunol* 3, 169-176, doi:10.1038/nri1004 (2003).
688 70. Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune
689 responses. *Nat Immunol* 5, 987-995, doi:10.1038/ni1112 (2004).
690 71. Lemaitre, B. & Hoffmann, J. The host defense of Drosophila melanogaster. *Annu Rev*
691 *Immunol* 25, 697-743, doi:10.1146/annurev.immunol.25.022106.141615 (2007).
692 72. Mottes, F., Villa, C., Osella, M. & Caselle, M. The impact of whole genome duplications
693 on the human gene regulatory networks. *PLoS Comput Biol* 17, e1009638,
694 doi:10.1371/journal.pcbi.1009638 (2021).
695 73. Holland, L. Z. & Ocampo Daza, D. A new look at an old question: when did the second
696 whole genome duplication occur in vertebrate evolution? *Genome Biol* 19, 209,
697 doi:10.1186/s13059-018-1592-0 (2018).
698 74. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in
699 complete eukaryotic genomes. *Genome Biol* 5, R7, doi:10.1186/gb-2004-5-2-r7 (2004).

700 75. Cunha da Cruz, J., Machado de Castro, A. & Camporese Servulo, E. F. World market and
701 biotechnological production of itaconic acid. *3 Biotech* 8, 138, doi:10.1007/s13205-018-
702 1151-0 (2018).

703 76. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local
704 alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2
705 (1990).

706 77. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
707 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780,
708 doi:10.1093/molbev/mst010 (2013).

709 78. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and
710 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol*
711 *Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

712 79. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S.
713 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**,
714 587-589, doi:10.1038/nmeth.4285 (2017).

715 80. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:
716 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522,
717 doi:10.1093/molbev/msx281 (2018).

718 81. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
719 tree display and annotation. *Nucleic Acids Res* **49**, W293-W296,
720 doi:10.1093/nar/gkab301 (2021).

721 82. Almagro Armenteros, J. J. *et al.* Detecting sequence signals in targeting peptides using
722 deep learning. *Life Sci Alliance* **2**, doi:10.26508/lsa.201900429 (2019).

723 83. Rangwala, S. H. *et al.* Accessing NCBI data using the NCBI Sequence Viewer and
724 Genome Data Viewer (GDV). *Genome Res* **31**, 159-169, doi:10.1101/gr.266932.120
725 (2021).
726 84. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**,
727 D412-D419, doi:10.1093/nar/gkaa913 (2021).
728 85. Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology
729 by domain architecture. *Genome Res* **12**, 1619-1623, doi:10.1101/gr.278202 (2002).
730 86. Shi, X. *et al.* Combinatorial G x G x E CRISPR screening and functional analysis
731 highlights SLC25A39 in mitochondrial GSH transport. *bioRxiv*, 2021.2009.2022.461361,
732 doi:10.1101/2021.09.22.461361 (2021).
733 87. Du, Y. et al. Validation of housekeeping genes as internal controls for studying gene
734 expression during Pacific oyster (Crassostrea gigas) development by quantitative real-
735 time PCR. Fish Shellfish Immunol 34, 939-945, doi:10.1016/j.fsi.2012.12.007 (2013).
736 88. Zhang, Q. L. *et al.* Selection of reliable reference genes for normalization of quantitative
737 RT-PCR from different developmental stages and tissues in amphioxus. *Sci Rep* **6**,
738 37549, doi:10.1038/srep37549 (2016).

739

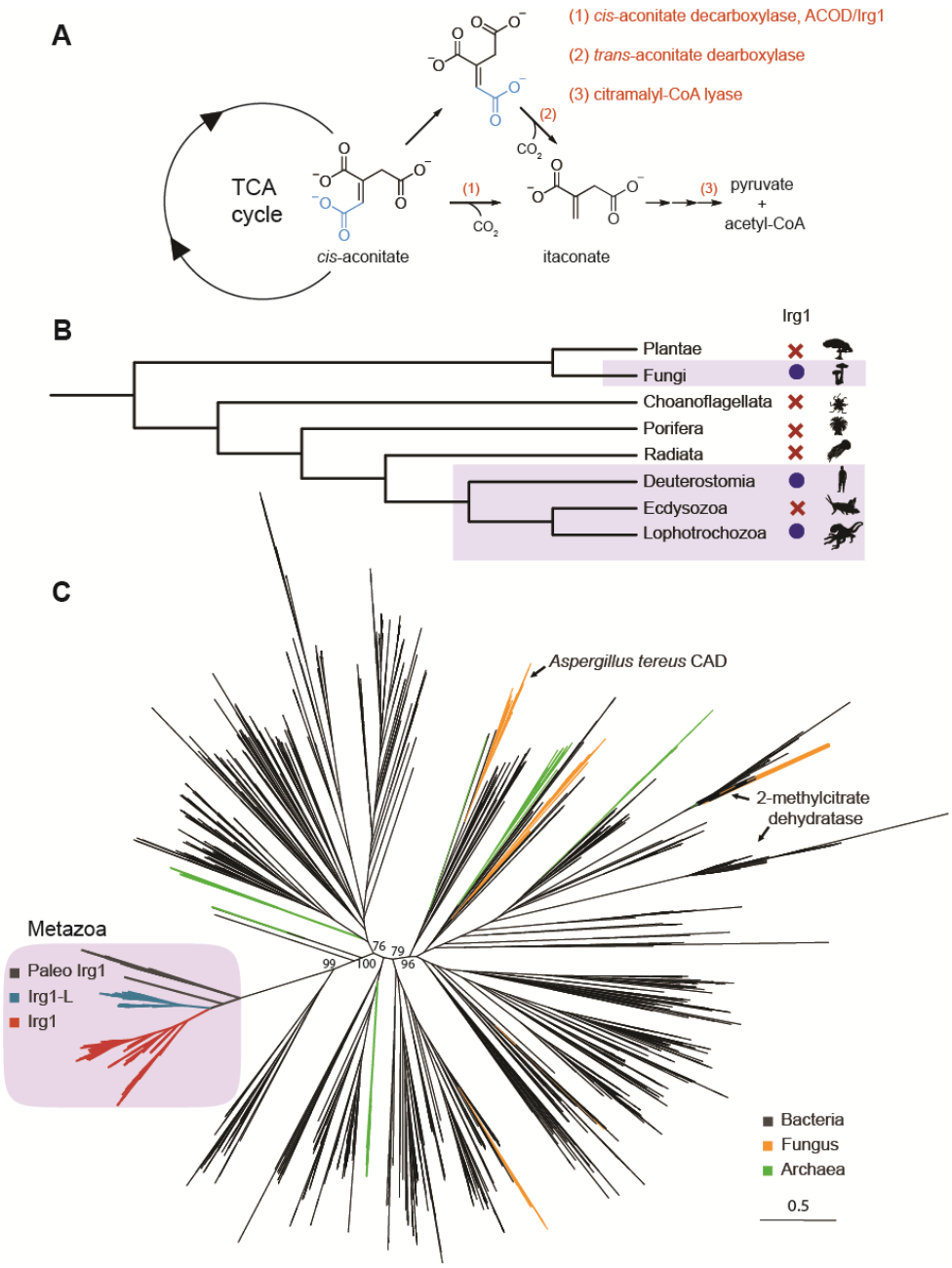
740

741

742

743

744




745

746 **Figure 1. Taxonomic distribution and molecular phylogeny of *cis*-Aconitate decarboxylase,**

747 **the enzyme producing itaconate, the innate immune metabolite.** (A) Itaconate biosynthesis

748 and degradation pathways in nature. Two itaconate biosynthesis pathways produce itaconate

749 from the TCA cycle intermediate *cis*-aconitate, through either the *cis*-aconitate decarboxylase

750 (CAD) enzyme or the trans-aconitate decarboxylase (TAD) enzyme. Itaconate degradation

751 involves a dedicated citramalyl-CoA lyase to generate pyruvate and acetyl-CoA. (B) Taxonomic
752 distribution of Irg1 in eukaryotes. Irg1 is present in Fungi, Lophotrochozoans, and
753 Deuterostomes but is absent in other metazoans. (C) Phylogenetic tree of Irg1 homologs in
754 nature. Bacterial sequences are colored in black, archeael in green and fungal in orange. High
755 confidence bootstrapping values are shown for the main branches separating the metazoan and
756 *A.tereus* CAD branches. Branches containing experimentally validated 2-methylcitrate
757 dehydratases are labeled. In the Metazoan branch, sequences were colored based on three
758 lineages, Irg1 (red), Irg1-like (Irg1L, cyan) and paleo-Irg1 (pIrg1, black).
759

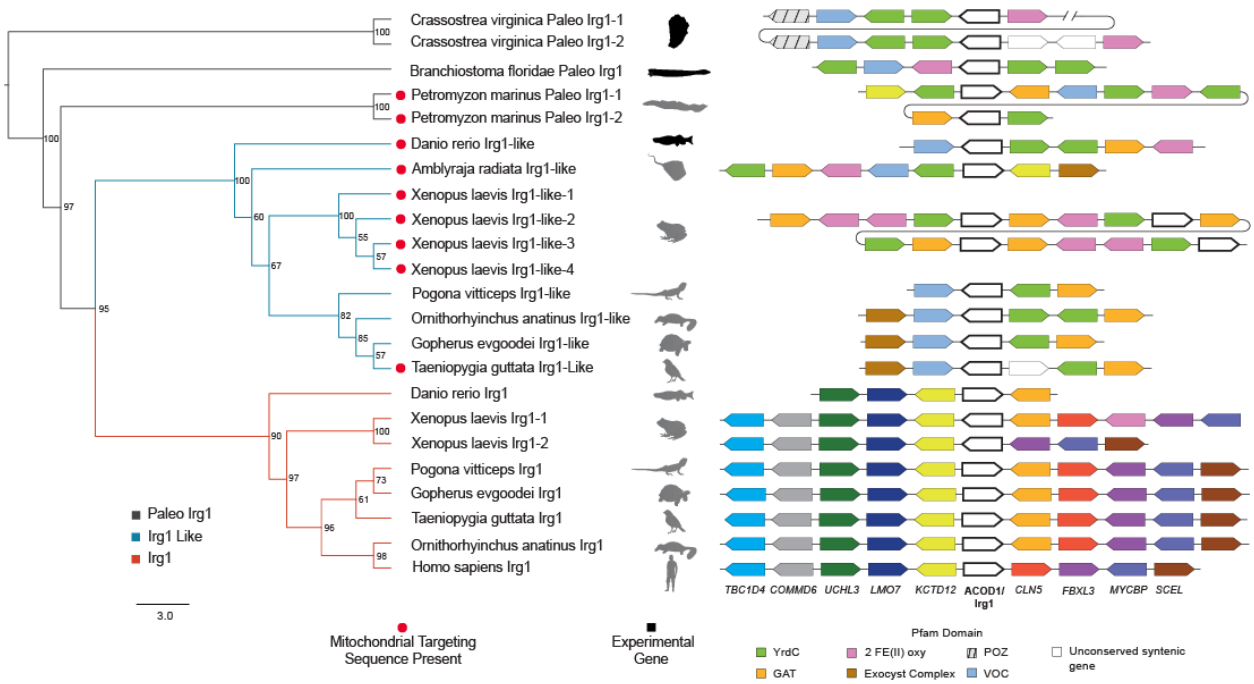
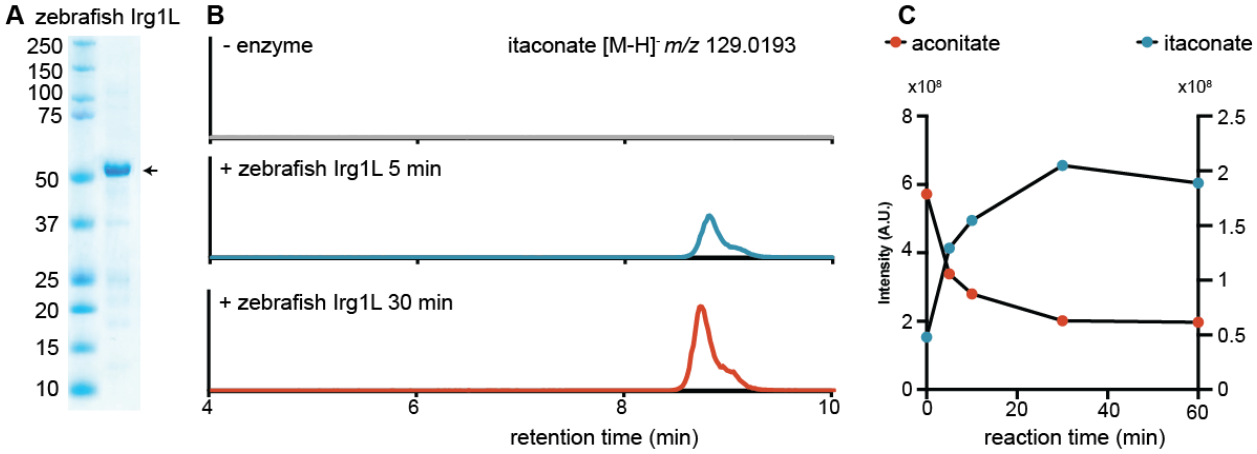



**Figure 2. Phylogenetic tree and synteny for representative Irg1 sequences in metazoa.** A subset of Irg1 sequence tree representing species from major vertebrate classes (grey). The tree branches are color-coded for the three lineages including paleo-Irg1 (pIrg1, black), Irg1-like (Irg1L, cyan) and Irg1 (red). The presence of the predicted Mitochondrial Targeting Sequence is highlighted by the red dots. Homologous genes in the genomic vicinity spanning the pIrg1, Irg1L and Irg1 are color-coded accordingly, 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.




774

775 **Figure 3. Zebrafish Irg1L is a *cis*-aconitate decarboxylase producing itaconate.** (A)

776 Recombinant His-tagged zebrafish Irg1L was expressed from *E. coli* and purified by affinity

777 purification. (B-C) LC-MS analysis confirms itaconate production from *cis*-aconitate (1mM) by

778 zebrafish Irg1L in a time-, enzyme-dependent manner.

779

780

781

782

783

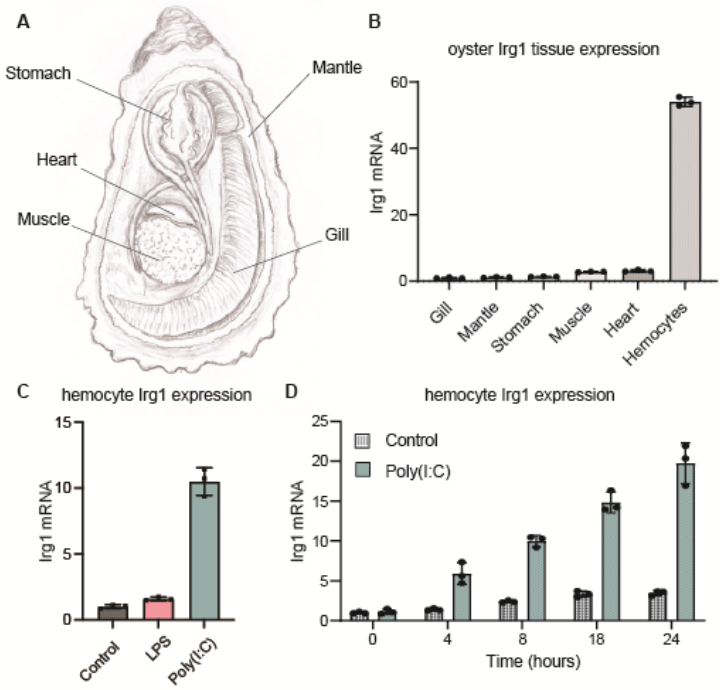
784

785

786

787

788




789

790 **Figure 4. The paleo-Irg1 in eastern oysters is implicated in innate immunity.** (A) oyster

791 anatomy. (B) Quantitative PCR showing oyster pIrg1 tissue expression. (C) The pIrg1 expression

792 in the oyster hemocytes is strongly induced by Poly(I:C) (10 ug/mL) but not LPS (1 ug/mL)

793 stimulation for 8 hrs shown by qPCR. (D) The time course of Irg1 induction in the oyster

794 hemocytes by Poly(I:C) (10 ug/ml) stimulation. Representative results are shown as mean ± s.d.

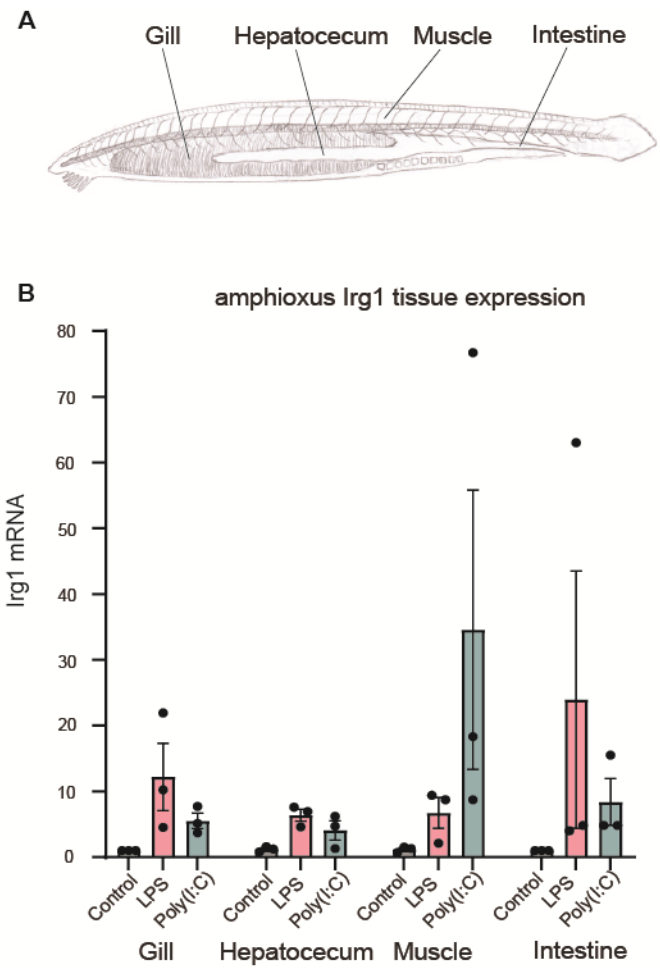
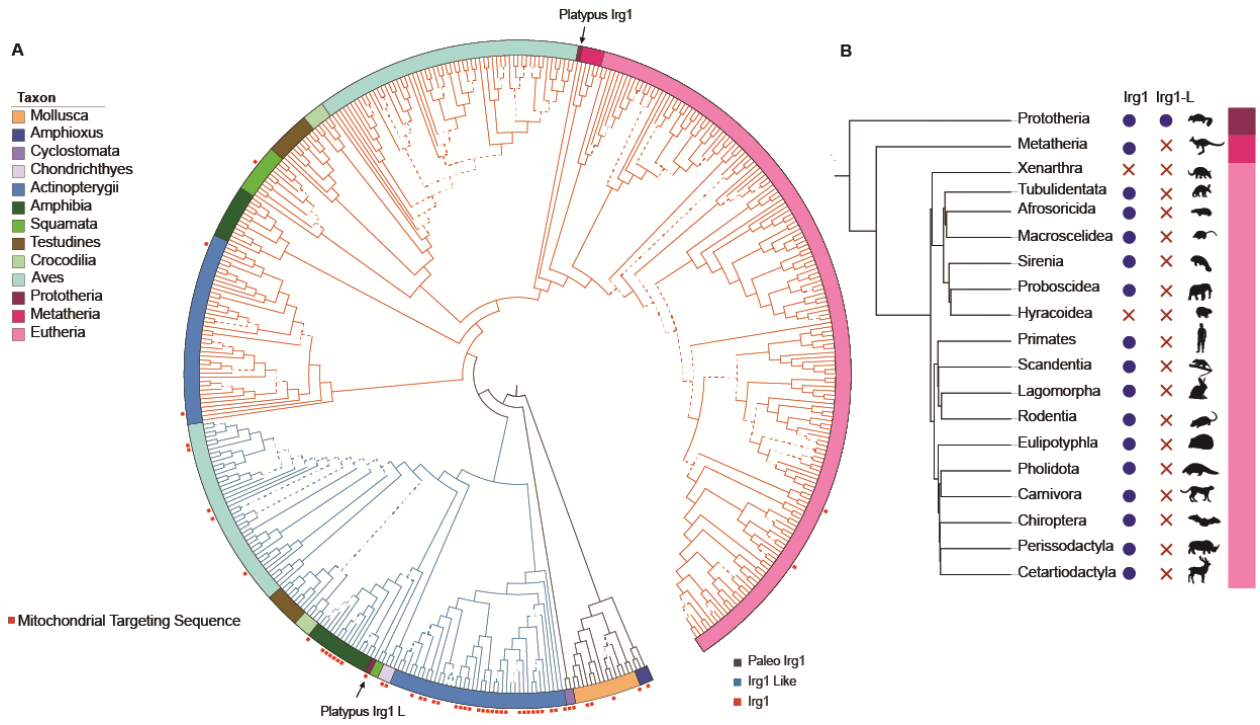
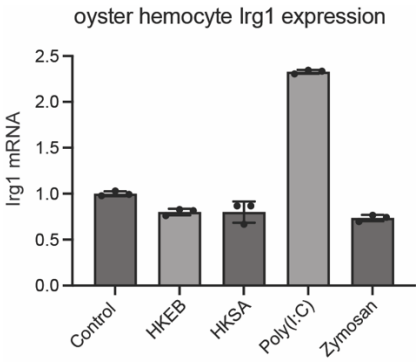



**Figure 5. The paleo-Irg1 in amphioxus is implicated in innate immunity.** (A) amphioxus anatomy. (B) Quantitative PCR from amphioxus tissues showing that pIrg1 is barely expressed under unstimulated condition, but can be strongly induced by both LPS (50 ug/mL) and Poly (I:C) (1 mg/mL) upon oral injection for 24 hrs. Each data point represents tissue from one animal shown as mean ± s.e.m. The large error bar is likely due to the low and variable mRNA level in the controls and individual animals' variability during stimulation.




**Supplemental figure 1. Phylogenetic tree and taxonomic distribution of pIrg1/Irg1L/Irg1 sequences in metazoa.** (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 for which the sequences belong to are color-coded on the outer rim of the tree. The sequences with predicted Mitochondrial Targeting Sequence were highlighted by red dots. (B) The presence of Irg1 and Irg1L in major mammalian taxa.




**Supplemental figure 2. PIrg1 induction in the eastern oyster hemocytes by innate immune stimuli.** Quantitative PCR showing the Irg1 mRNA level 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).

| 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' |
| 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' |

839

840 **Table 1**. Primer sequences used for RT-qPCR in *C. virginica* and *B. floridae* pIrg1 transcription

841 assays.

842

843

844

845

846

847

848

849

850