Coupled bio-physical global ocean model and molecular genetic analyses identify multiple introductions of cryptogenic species

Michael N Dawson¹, Alex Sen Gupta², and Matthew H. England²

¹ Biological, Earth, and Environmental Sciences
² Centre for Environmental Modelling and Prediction, School of Mathematics
University of New South Wales, Sydney, NSW 2052, Australia

Corresponding author: Michael N Dawson
Evolution and Ecology, Storer Hall, UC Davis, One Shields Avenue, Davis, CA 95616, USA
Email: mndawson@ucdavis.edu  Tel.: 530-752-2937  Fax: 530-752-1449

Manuscript information: 21 text pages, 4 figures, 1 table.

Words in abstract: 207  Characters in paper: 38862 (text only; table legend 411)
59509 (with equations, figures, table)

Abbreviations: COD, cumulative occurrence distribution. COI, mitochondrial cytochrome c oxidase subunit I. NIS, non-indigenous species. nrDNA, nuclear ribosomal DNA. POCM, Parallel Ocean Climate Model.

DNA sequence data deposited in GenBank: COI AY903067-AY903213, nrDNA AY935202-AY935218.
The anthropogenic introduction of exotic species is one of the greatest modern threats to marine biodiversity. Yet exotic species introductions remain difficult to predict and are easily misunderstood because knowledge of natural dispersal patterns, species diversity, and biogeography is often insufficient to distinguish between a broadly dispersed natural population and an exotic one. Here we compare a global molecular phylogeny of a representative marine meroplanktonic taxon, the moon-jellyfish *Aurelia*, with natural dispersion patterns predicted by a global bio-physical ocean model. Despite assumed high dispersal ability the phylogeny reveals many cryptic species and predominantly regional structure with one notable exception: the globally distributed *Aurelia* sp.1 which, molecular data suggest, may occasionally traverse the Pacific unaided. This is refuted by the ocean model, which shows much more limited dispersion and patterns of distribution broadly consistent with modern biogeographic zones, thus identifying multiple introductions world-wide of this cryptogenic species. This approach also supports existing evidence that (i) the occurrence in Hawai‘i of *Aurelia* sp.4 and other native Indo-West Pacific species with similar life-histories is most likely due to anthropogenic translocation, and (ii) there may be a route for rare natural colonization of northeast North America by the European marine snail *Littorina littorea* whose status as endemic or exotic is unclear.

Until recently (1) marine species introductions were of limited concern because many marine plankton were assumed to have naturally broad, even global, distributions (2-4). However, as marine molecular genetics and physical oceanography have increasingly revealed biotic and physical discontinuities in an ecologically heterogeneous environment (4-6), anthropogenic introduction of non-indigenous species (NIS) has been recognized as a major threat to native biodiversity (7, 8) with the potential to alter ecosystems (9, 10), displace endemic species, and cost millions of dollars in damage and preventative control (11). Yet the threat is still poorly
understood (7) and species introductions remain generally unpredictable (12) because
knowledge of natural dispersal patterns (4, 6), species diversity (13), and biogeography (14) is
often insufficient to identify non-indigenous species or their sources. Recent surveys classified
36%-47% of NIS, which may constitute 13%-23% of species in international ports, as
“cryptogenic” (i.e. neither clearly native or introduced) or “cosmopolitan” (i.e. distributed
globally; 8, 15).

Molecular genetic techniques have been heralded as a powerful tool for taxonomic
verification, differentiating cryptogenic taxa, identifying source populations and vectors, and
assessing the extent and impacts of invasions (16, 17). However, this approach to NIS relies
on the questionable assumption that introduced populations are characterized by massively
reduced genetic diversity and no unique genotypes (17) which will generate estimates of gene
flow dating only to the last few centuries. Ocean-going vessels typically visit numerous ports
multiple times carrying sufficient numbers of planktonic, colonial, or aggregating fouling
invertebrates for successful reproduction and invasion (18). This has the potential to increase
genetic diversity in the invaded range creating, with sampling error, patterns indicative of
endemicity predating human trans-oceanic voyages. As a result, molecular genetic analyses
alone may be insufficient to assess whether a geographic occurrence represents a species
introduction. Here, we demonstrate that simulating natural dispersion over the time-scales of
human travel aids identification of NIS and, at the same time, elucidates taxonomy, patterns of
marine biodiversity, and biogeography.

Materials and methods

Molecular analyses. Tissues from 78 Aurelia medusae were sampled and preserved in 70-90%
ethanol or salt-saturated dimethyl sulfoxide (19). Mitochondrial cytochrome c oxidase subunit I
(COI) was amplified using primers LCOjf (20) and HCO (21) and PCR conditions described in (20). Nuclear ribosomal DNA (nrDNA) was amplified from 14 of these medusae, representing each major COI clade (in duplicate when possible), using primers jfITS1 and 28S-2R and PCR conditions described in (22). COI amplicons were purified using Qiagen PCR clean-up columns, rDNA amplicons cloned using TOPO TA (Invitrogen) and purified using Pharmacia’s Flexiprep kit. Purified amplicons were labelled with BigDye and sequenced on ABI 377 automated sequencers according to Applied Biosystems protocols at the University of New South Wales’ Ramaciotti Centre. Electropherograms were checked visually and misreads corrected. Homologous sequences from prior studies (20, 22-25) were appended to each dataset then COI sequences were aligned on the basis of the amino acid translations whereas nrDNA sequences were aligned using several gap-opening:gap-extension weighting schemes in ClustalX (26) and then corrected by eye. Positions with missing data were excluded from subsequent analyses. The total dataset comprises 240 medusae from 77 sites world-wide of which 174 were sequenced for COI and 155 for nrDNA (Supporting Information online). Due to the large number of specimens, the phylogenetic analyses reported here use a subset of sequences for which both COI and nrDNA data were available and represented major reciprocally monophyletic clades found in prior studies and/or preliminary maximum parsimony analyses of the entire dataset. These clades were assigned species status (20, 22-25) if recovered in analyses of mitochondrial and nuclear markers and if separated by sequence differences comparable to or greater than those seen between the traditionally recognized morpho-species *A. aurita* and *A. limbata* (i.e. circa 10% in COI and ITS1 [23]).

Gene trees were reconstructed using maximum likelihood in PAUP*4.0b10 (27) using appropriate models of molecular evolution (COI, GTR+G; nrDNA, TrNef+I+G) identified by
MODELTEST (28). The robustness of the *Aurelia* sp. 1 sister taxon relationship identified by phylogenetic analyses was tested against all other possible pairings using the Kishino-Hasegawa Test in PAUP*4.0b10 employing RELL bootstrap (100,000 replicates) and a one-tailed test (29). Bootstrap analyses consisted of 1,000 realizations using the same maximum likelihood models and also unweighted maximum parsimony (10,000 realizations) including gapped positions in nrDNA. Additional COI data describing *Aurelia* sp.1 were used to reconstruct an haplotype network using unweighted parsimony in PAUP*4.0b10. Haplotype and nucleotide diversity and population subdivision ($\phi_{ST}$) were calculated in Arlequin 2.0 (30).

**Ocean bio-physical modelling.** To investigate the limits of potential natural dispersion of species of *Aurelia* over multi-century timescales we developed a global two dimensional Lagrangian model incorporating representative biological life-history characteristics of *Aurelia*. These life-history characteristics include a small, probably perennial, benthic polyp that reproduces asexually to produce other benthic polyps and free-living planktonic medusae; the medusae reproduce sexually, producing a planula larva that is brooded for a short period by the female then released into the water column where it spends probably <1 week before settling on the benthos where it metamorphoses into a polyp. Thus, the medusa is the main dispersive phase; medusae usually live < 6 months in nature although medusae > 1 year have been recorded; they may live up to 2 years in captivity (31, 32; Supporting Information online).

The circulation model is driven by monthly averaged advection fields from a 20-year integration (1979–1998) of the Parallel Ocean Climate Model (POCM; 33-35). The current version (POCM 4C) simulates ocean circulation in the latitudinal domain 68°N to 75°S on a Mercator grid with latitude-longitude resolution of $0.4^\circ \cos(\phi) \times 0.4^\circ$, ($\phi$ is latitude), resulting in an average grid size of 0.25° in latitude. Surface forcing of the POCM consists of 20 years
of daily varying momentum, heat and freshwater fluxes derived from the European Centre for Mid-Range Weather Forecasting (ECMWF) model reanalysis (1979 - 1993) followed by operational fields from the ECMWF (1994–1998; 35). The POCM accurately represents surface circulation patterns (34) and interior water-mass pathways (36). We take the ocean to be in a steady-state seasonal cycle derived from a 20-year mean of the POCM simulation, thus ignoring interannual circulation variability such as that due to El Niño Southern Oscillation (ENSO) or millennial scale fluctuations in the ocean’s thermohaline circulation. This simplification reasonably reflects the relative stability of the Earth’s climate during the current interglacial period (37). Changes in ocean circulation over at least the past seven millenia have been primarily limited to subtle shifts in the location and intensity of the subtropical gyres and to natural variability associated with phenomena such as ENSO and the North Atlantic Oscillation. The sensitivity of the simulated advection pathways to this assumption of internal oceanic variability is addressed below.

Each off-line model experiment is based on the release of ca. 20,000 particles from known Aurelia sp. 1 zones of occurrence with particle positions \( x_t \), at time \( t \) given by

\[
x_{t+1} = x_t + U(x_t) \Delta t + R_n \frac{\sqrt{2K_i \Delta t}}{\sqrt{\Delta t}} + C_{mix}.
\]

Here \( \Delta t \) represents the model time-step (6 hours) and \( U \) is the flow velocity interpolated to the position of the particle. Particles are forced to remain active by ignoring any component of \( U \) that would cause the particle to become grounded. The POCM resolves ocean dynamical processes at the mesoscale (tens of kilometres) and above (e.g. gyres, western boundary currents, deep ocean jets). This degree of resolution permits a certain amount of eddy mixing in prognostic mode. However, the aliasing resulting from the temporal averaging of \( (u, v) \) will significantly reduce any eddy-mixing in the off-line model (36). To account for this missing
eddy advection, a constant horizontal diffusivity \( (K_h = 1000 \text{ m}^2 \text{s}^{-1}) \) is incorporated into the Lagrangian dispersal model. The diffusivity is formulated as a random-walk term, where \( R_N \) is a normally distributed random number (mean = 0, standard deviation = 1). We tested the effects of halved, doubled, and five-fold greater diffusivities, which encompasses the values used by a large number of global climate models. None of our conclusions regarding natural Aurelia dispersal were altered (Supplementary Information online). Lower values, such as \( K_h = 100 \text{ m}^2 \text{s}^{-1} \) used to investigate Lagrangian krill transport in the Southern Ocean (38) would only further limit dispersal potential. Our experimental philosophy was to run with generous diffusivity to give the particles’ simulated range extent as an upper bound on natural dispersal.

To improve particle migration through ‘island hopping’ the POCM land mask was supplemented with the 5-minute ETOPO5 bathymetry dataset. Ocean gridpoints in the POCM that contained islands in ETOPO5, were assigned a fractional number indicating the percentage of the POCM grid containing land. Land was considered to be any area with an ocean depth less than 10m. During the simulations, for each successive re-release, particles were released from these ‘island’ gridboxes in proportion to the number of particles having impacted on that region during the previous year, scaled by the land fraction. An investigation of the 2-minute ETOPO2 dataset showed no additional island sites at positions that would aid particle stepping.

Release of the Aurelia particles occurs over an entire year, but is biased towards summer months, corresponding to normal patterns of abundance of Aurelia medusae (32). Each particle advects through the surface 75 m of the ocean for one ‘lifetime’ (i.e. for up to 365 days corresponding to a reasonable approximation of the upper life span of Aurelia; 31, 32) after which time it ‘dies’. For each model grid box, a cumulative total of the number of particles in that grid-box per time-step is calculated, providing a cumulative occurrence
distribution (COD) representing the range of possible positions (and time spent at each location) at which a medusae may have released planulae larvae potentially resulting in colonization of the area by the benthic polyp stage. Once all particles have exceeded their lifetimes a new release of particles is initiated with new coastal release locations calculated as a function of the coastal COD values. This process is repeated, at the completion of each set of lifetimes, using the newly generated COD.

To represent unresolved coastal flows, new release locations (determined by the COD) are recalculated, before each successive re-release, by applying an effective diffusion along the coastline as a result of estimated tidal currents ($C_{mix}$). An additional constant background alongshore current (1.0 m s$^{-1}$) is also added to the modelled tidal current to account for any non-tidal flows (e.g. coastal trapped waves, transient coastal currents, and shelf waves). The tidal current speeds are derived from a global inverse tidal model (TPXO.6; 39, 40; Supplementary Information online). A temperature dependent modification is also made before each new release, with the COD being reduced by a factor $\sigma(x,y)$, at temperatures ($T(x,y)$) above and below the annually averaged temperature maximum ($T_{max} = 24.5$ °C) and minimum ($T_{min} = 14$ °C), respectively, of known *Aurelia* sp.1 habitats, such that

$$\sigma(x,y) = \frac{1}{2\Delta T(x,y)}$$

where

$$\Delta T(x,y) = \begin{cases} T_{min} - T(x,y), & \text{for } T < T_{min} \\ T(x,y) - T_{max}, & \text{for } T > T_{max} \end{cases}$$

The exponentially decreasing rate of survivorship with increasing deviation ($\Delta T$) from temperatures normally experienced is a general approximation derived from data describing temperature-dependent mortality in the scyphozoan jellyfish *Mastigias* (41). Mortality rates in *Mastigias* increase slowly with increasing deviation above average temperatures to a point at which mortality rate begins to increase rapidly, a pattern that is broadly consistent with data on other marine invertebrates, such as corals (e.g. 42) and gastropods (43) that show
exacerbated effects at elevated temperatures. The temperature-mortality relationship used is a reasonable general approximation of the observed relationship in *Mastigias* allowing for variation between populations (41, 43), species, and interactions with other environmental variables (43).

Distributions reach an approximate steady-state within a century of model simulation time. This is expected as the life-history of *Aurelia* limits the geographic extent over which the medusae can advect or mix, even considering possible routes of migration via island stepping stones and coastal-zone diffusion. As a result only the Japan release experiment was integrated for a full 10,000 years; all other experiments were integrated for 1,000 years, well beyond the simulation’s quasi-steady state.

Transient interannual change in ocean circulation is associated with El Niño events but the impact of such ocean variability on medusae advection is ignored in the current experiments which use monthly advection fields averaged from 20 years of the POCM integration. To investigate the possible impact of interannual regional circulation variability on the simulated advection pathways, additional sensitivity experiments were carried out wherein the medusae were advected under perpetual El Niño and La Niña scenarios. These artificial circulation scenarios represent an upper limit for the effects of interannual changes in Pacific Ocean conditions, providing a stringent test of the robustness of our findings with respect to natural dispersion of *Aurelia*. Although shifts occurred in the final distributions and maximum particle migrations consistent with expectations of ENSO-altered circulation patterns (Supporting Information online) the results with respect to the absence of translocation between the *Aurelia* sites were robust.

**Results and Discussion**
The global phylogeny of *Aurelia* reveals at least 16 phylogenetic, i.e. 13 cryptic, species (Fig. 1, see also Supporting Information online) most of which appear to be regionally restricted. Several species, however, have disjunct distributions, which may be a characteristic of introduced species. These include *Aurelia* sp.4 which occurs in the Western Pacific and in Pearl Harbor typical of introductions dating to the Second World War (15), *Aurelia aurita* which is endemic to the North Atlantic and disjunct in the Black Sea like the introduced ctenophore *Mnemiopsis* (9), and *Aurelia* sp.8 which has a “Lessepsian” distribution, i.e. occurs on both sides of the Suez Canal, typical of other introduced species including the scyphozoan jellyfish *Rhopilema nomadica* (44). Most remarkable is *Aurelia* sp.1 (a cryptic species of *A. aurita*), which occurs in major warm-temperate regions around the globe (Figs. 1, 3). The sister-taxon relationship between *Aurelia* sp.1, *A. limbata*, and *Aurelia* sp.10 (see Supporting Information online) indicates that *Aurelia* sp.1 is endemic to the western North Pacific and, therefore, dispersed globally from Japan. The anomalously broad distribution of *Aurelia* sp.1, however, is suggestive of anthropogenic introduction, since COI shows reduced molecular diversity in Australia and California compared to Japan (Fig. 2) and estimates of divergence times using a scyphozoan COI ‘molecular clock’ do not exclude the modern day (Table 1). However, average estimated divergence times precede global shipping by thousands or tens of thousands of years (Table 1) suggesting rare, but natural, long-distance dispersal on evolutionary time-scales, consistent with the hypothesis that *Aurelia* sp.1 is naturally globally distributed (2, 24) and that 95% confidence intervals encompassing zero divergence times simply reflect ongoing gene flow.

Anthropogenic range expansions are distinguished by the fact that they exceed an organism’s natural dispersal ability. Our model translocation results (Figs. 3, 4) demonstrate
that *Aurelia* can occupy ranges of thousands of kilometres and presumably disperse over these geographic distances on evolutionary time-scales, consistent with interpolated ranges of most phylogenetic species of *Aurelia* (Fig. 1). However, natural dispersal and mixing occur only within limited geographic regions. For example, in the North Pacific, *Aurelia* spread from Japan northwards and eastwards in the Kuroshio Current, as well as mixing locally in the Yellow Sea, the East China Sea, and the Sea of Japan. Dispersal beyond this is limited because the simulated 1-year life-span of medusae is much less than the time required (c. 5-10 years) to traverse the North Pacific basin by advection alone. There are no temperate island stepping-stones in the North Pacific that might facilitate multi-generational trans-oceanic dispersal of *Aurelia* sp. 1, and the Aleutians are too far poleward and hence too cold to facilitate coastal trans-Pacific migration of *Aurelia* sp. 1, even with along-shore mixing set rather high. The model simulates no natural dispersion of any propagules from Australia to North America, or vice versa, under modern circulation fields (i.e. to circa seven millenia before present) because possible advection pathways are well beyond the modelled 1-year life-span of *Aurelia* medusae (even using coastal and island stepping stones) and/or are prohibited by regions of inhospitable water temperatures. This is at odds with the minimum of two successful colonization events, and implicitly many more unsuccessful dispersal events, of *Aurelia* sp. 1 indicated by genetic data (Fig. 2) on comparable or only slightly longer time-scales (Table 1) and we interpret the discrepancy as evidence of multiple human-mediated translocations of *Aurelia* between these two sites. This interpretation is consistent with the inferred appearance of *Aurelia* in many global locations coincident with periods of heavy Pacific vessel traffic (8, 15, 48). For example, *Aurelia coerulae* von Lendenfeld 1884 (presumed synonymous with *A. japonica* Kishinouye 1891 and *Aurelia* sp.1) was first described from Port Jackson, Sydney, Australia,
following a major increase in shipping from the northwest Pacific (8), and the first occurrences
of *Aurelia* sp. 1 in California are poorly circumscribed to the late 1900s (48) consistent with
20th Century increases in trans-Pacific shipping (8). Thus, multiple introductions is a more
parsimonious interpretation of all available evidence than natural dispersal predating or
continuing during the Holocene and the modern day. According to the existing model results,
the genetic diversity of *Aurelia* sp. 1 in Europe (24) must also result from multiple
introductions.

The discrepancy between the implications of molecular and ocean modelling analyses
highlights two genetic effects with important consequences for our understanding of global
NIS. First, estimates of local diversity in NIS populations are inflated by multiple
introductions, thus masking the expected low-diversity genetic signature of introduced species.
Second, estimates of the timing or duration of gene flow can be inflated by multiple
introductions, especially if dispersal via human vectors favours differential establishment of
alleles rare in the natural range (18). Both erode strong genetic signals expected in NIS (17)
giving the false appearance of a naturally dispersed endemic population.

The ocean model can also address other anthropogenic introductions. For example,
*Aurelia* sp. 4 is endemic to eastern Borneo and Palau and also occurs in Hawai’i (Fig. 1). The
ocean model shows that *Aurelia*-like particles released off Borneo are advected eastwards to
Palau by an eddy in the Celebes Sea and via the Equatorial Counter Current but that none
disperse to Hawai’i even on multi-century time-scales (Fig. 4). This result is robust even under
the most extreme (to the point of being unrealistic) dispersal scenario of a perpetual El Niño
circulation field, demonstrating that, for a species with the dispersal characteristics modeled for
*Aurelia* sp.4, there is no available ocean pathway that naturally connects these zones of
occurrence on these time-scales (see also Supporting Information online). The range-limit indicated in the model (Fig. 4) is consistent with preliminary collections that found *Aurelia* sp. 11 (not sp.4) in the Marshall Islands and, more generally, with Indo-West Pacific biogeography wherein many species that occur in the Indo-West Pacific centre of marine biodiversity do not penetrate into the central Pacific. For example, there are ~2500 species of shorefishes in the Philippines, ~1300 in Palau, ~850 in the Marshall Islands, and only ~550 in Hawaii (49, 50); similar patterns are evident in scyphozoan jellyfishes (51, 52), corals (53), and many other taxa. Concomitantly, historical records reveal that *Aurelia* is not native to Hawai’i. Despite surveys of scyphozoans in the early 1900s (54, 55) the genus *Aurelia* was first reported in Hawaii in Pearl Harbor in 1953 (56) following particularly high WWII naval traffic with the western Pacific (J. Carlton, pers. comm.). Thus, the reciprocity of historical, biogeographic, molecular, and model datasets helps validate the model’s solution which now provides additional support to the existing body of evidence (e.g. 15, 57) that a substantial number of endemic Indo-West Pacific species which also occur in Hawai’i but have limited natural dispersal ability, including other scyphozoans with life-histories like *Aurelia* sp.4 (58-60), are most probably introduced.

Like evidence of disjunction, evidence of connectivity in the ocean model is also important for interpreting data on species introductions. The model identifies a possible rare route of natural dispersal across the North Atlantic, via the northern limb of the sub-polar gyre (Fig. 3), potentially supporting recent genetic evidence that *L. littorea* populations in northeast North America are endemic resulting from natural range expansion predating Viking expeditions (17). The projected pathway is sensitive to the choice of temperature dependence and model boundary conditions, but the route identified by the model should, if anything, be
stronger for the boreal *L. littorea* than those shown in Fig. 3 for the warm-temperate *Aurelia*. This rare dispersal route is consistent with evidence of different varieties of *A. aurita* in Europe and North America (2, 23) and of many east-west amphi-Atlantic species, particularly boreal and Arctic-boreal molluscs and fishes (61). However, the conclusion of natural range expansion based on genetic data (17) are not indisputable and it remains to be demonstrated that observed patterns and levels of genetic diversity are explained better by rare dispersal of *L. littorea* than by multiple introductions.

The patterns of connectivity and disjunction identified by the genetic data and ocean trajectory model are also broadly consistent with other previously recognized biogeographic patterns (61). For example, the highest particle densities in the Australian simulations are concordant with the Flindersian Province, the eastern boundary of which is marked by an abrupt decrease in transport around Cape Howe (the southeastern-most tip of mainland Australia), although some propagules do enter the eastern transition zone which peters out around Brisbane (62, 63). The Australian simulation also shows some connectivity between west-coast and east-coast warm temperate faunas, via an eastward flowing extension of the Leeuwin Current, and connectivity between eastern Australia and New Zealand (particularly the North Island), via the southwestern limb of the sub-tropical gyre, consistent with the distribution of *Aurelia* sp. 7 and biogeography (61). These results suggest a strong role for hydrography in shaping modern patterns of species diversity (5, 20, 64), which is a salutary finding in light of the growing emphasis, albeit debated and based largely on studies of fishes, on the role of behavior in generating or maintaining geographic structure in marine taxa (e.g. see 6; 65-67). Ultimately, extrinsic (e.g. currents) and intrinsic (e.g. behavior, life-history) processes interact, with probably different relative effects on different larvae (6).
Consequently, as geographic isolation of marine taxa gains renewed attention (4-6), we need to develop tools that improve our ability to understand the interacting dispersive and isolating influences of ocean currents and animal behavior and their contributions to spatial patterns of genetic variation and evolution in marine taxa. Their influences may not always be obvious. Our global ocean model has demonstrated that even marine species with long planktonic stages, such as *Aurelia*, can be regionally restricted due to natural oceanographic patterns, thus having higher than expected geographic structure and species diversity.

Therefore, relative to prior expectations, marine populations may be more susceptible to be threatened by, or to become, invasive species due to anthropogenic introductions that exceed limited natural dispersal ability.

M. Baird, J. Carlton, B. Holland, L. Martin, and two anonymous reviewers provided valuable comment. T. Bolton, J. Gardner, H. Ishii, H. Miyake, M. Omori, P. Petersen, S. Uye, and S. Wilcox contributed samples. We were supported by a Vice-Chancellor’s Post-Doctoral Fellowship from University of New South Wales (MND), an International Postgraduate Research Scholarship (ASG), and an Australian Research Council Large Grant (MHE).

16


30. Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver. 2.000: a software for population genetics data analysis* (Genetics and Biometry Laboratory, University of Geneva).


Supporting Information online

Figures S1 (Aurelia life history), S2 (El Niño simulations), S3 (coastal dispersion), S4 (ocean diffusivity), S5 (island stepping stones), S6 (ITS1 gene tree), S7 (COI gene tree), S8 (Map of names).

Tables S1 (specimen list), S2 (Kishino-Hasegawa and Wilcoxon signed ranks test results)
Fig. 1. Molecular phylogeny of *Aurelia* reveals many cryptic species with predominantly limited geographic distributions. Maximum likelihood nrDNA (-Ln = 7425) and COI (-Ln = 4725) gene trees. Bootstrap values >50% are shown above (maximum likelihood, 1000 realizations, excluding gapped positions in nrDNA) or below (unweighted maximum parsimony, 10000 realizations, including gapped positions in nrDNA) each branch. Scale bars represent 0.025 substitutions/site. Trees were rooted with sequences from *Cyanea* spp. (and also *Phacellophora* COI). *Calculated excluding outgroup taxa; support for other nodes increased 3-10% over values shown, or remained at 100%.

Fig. 2. Network of the most parsimonious relationship between the 20 COI haplotypes sequenced from *Aurelia* sp.1. Each circle represents a different haplotype; the area of each circle or segment is proportional to the frequency with which that haplotype was observed (largest to smallest circles: 30, 12, 6, 5, 3, 2, 1). The color of each segment indicates the geographic origin of that fraction of the samples. Each branch of unit length represents a one-nucleotide mutation. Small squares indicate a haplotype that does, or did, exist but which we did not sample. Genetic diversity (mean ± sd) is high in Japan (h = 0.87 ± 0.06, π = 0.0063 +/- 0.0037, n = 26) and low elsewhere (California: h = 0.53 +/- 0.14, π = 0.0020 +/- 0.0015, n = 16; Australia: h = 0.66 +/- 0.08, π = 0.0048 +/- 0.0029, n = 37). In subsequent analyses, eastern and western Australia (EA and WA) are treated as one unit due to the small WA sample (n = 6), because all WA haplotypes are found in EA so there is no significant difference between regions (ΦST = 0, p = 0.97), and because ocean modelling indicates EA and WA may be well connected on evolutionary time-scales (Fig. 3). Most California sequences taken from (23) were cloned, whereas new sequences for this study were direct sequenced. Consequently, in contrast to other unique haplotypes, the three unique California haplotypes may result from PCR error made unambiguous by sequencing cloned amplicons.

Fig. 3. Final year cumulative occurrence distributions (COD) – coloured marine areas and scale - for releases in the five primary zones of occurrence of *Aurelia* sp.1 (adjacent red land areas; main, east Australia release; inset, west Australia release). The COD around Japan is derived from a 10,000 year simulation, the others from 1,000 year simulations; all have reached a quasi-stable state. Black and red
contours represent estimated maximum extent of particles based on samples taken over the integration period for open and closed North Atlantic boundary, respectively. COD is the sum over all timesteps of the number of particles in a particular gridbox for the lifespan (1yr) of the particles. Grey shading (see scale bar) indicates the proportional effect of temperature on survivorship of medusae prior to reproduction and can be interpreted as an index of establishment probability or reproductive viability of a population, integrated over the medusa and polyp phases, consistent with evidence of temperature effects in *Aurelia* (24) and with species introductions that occur predominantly along lines of similar latitude (45). Symbols show the distribution of known phylogenetic species of *Aurelia* based on COI and nrDNA (circles), 16S and partial-nrDNA (squares), or COI (triangle) sequence data (see Supporting Information online). Species are numbered as in Fig. 1, or represented by a letter code: A, *Aurelia aurita*; B, *Aurelia labiata*; CA, *Cyanea-Aurelia* hybrid (39); M, *Aurelia limbata*; R, *Aurelia "ARAB"* (39). An additional nrDNA haplotype “MCA” of unknown geographic origin has been documented (39, see Supporting Information online).

Fig. 4. Final year cumulative occurrence distribution for initial releases of *Aurelia* sp.4 from Borneo and the Philippines, i.e. within the region in which *Aurelia* sp.4 is endemic. Contours, shading, and symbols are as described in Fig. 3. Symbols for Palau are shifted southward to reveal the effect of the Equatorial Counter Current.
Table 1. Estimated divergence times (T; to nearest millenium) between Pacific populations of *Aurelia* sp.1. Divergence time and 95% CI estimated from net pairwise divergence (17) \( T = d/2\lambda \) where \( d = D_{xy} - (D_x + D_y)/2 \) and \( \lambda \) (point mutation rate/year) = 1.47x10\(^{-8}\) (20; see also 46, 47). \( D_{xy} \) = average number of pairwise differences between regions; \( D_x, D_y \) = average number of pairwise differences within regions.

<table>
<thead>
<tr>
<th></th>
<th>d</th>
<th>95% CI</th>
<th>T (years)</th>
<th>95% CI (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>California-Australia</td>
<td>0.036</td>
<td>±0.038</td>
<td>12,000</td>
<td>±13,000</td>
</tr>
<tr>
<td>California-Japan</td>
<td>0.431</td>
<td>±0.038</td>
<td>147,000</td>
<td>±13,000</td>
</tr>
<tr>
<td>Australia-Japan (20)</td>
<td>0.202</td>
<td>±0.028</td>
<td>69,000</td>
<td>±10,000</td>
</tr>
</tbody>
</table>