The small genetic world of *Seriatopora hystrix*

Stuart Kininmonth¹, Madeleine van Oppen¹, Sarah Castine², Lesa Peplow¹, Adrian Lutz¹

¹Australian Institute of Marine Science, PMB#3, Townsville, Queensland, 4810 Australia
²AIMS@JCU, DB17-063, James Cook University, Townsville, QLD 4811 Australia

E-mail: stuart@kininmonth.com.au

Received 12 August 2011; Accepted 15 September 2011; Published online 20 November 2011

**Abstract**

The exchange of genetic information among coral reefs, through the transport of larvae, is critical to the function of Australia’s Great Barrier Reef because it influences recruitment rates and resilience to disturbance. For many species the genetic composition is not homogeneous and is determined, in part, by the character of the complex dispersal pathways that connect the populations situated on each coral reef. One method of measuring these genetic connections is to examine the microsatellite composition of individual corals and then statistically compare populations across the region. We use these connection strengths, derived from a population similarity measure, to create complex networks to describe and analyse the genetic exchange of the brooding coral, *Seriatopora hystrix*. The network, based on determining the putative parental origin of individual coral colonies, involved sampling 2163 colonies from 47 collection sites and examining 10 microsatellites. A dispersal network was created from the genetic distance DLR values that measure the genetic similarity of each population (defined by the local sampling effort) to every other sampled population based on the microsatellite composition. Graph theory methods show that this network exhibited infrequent long distance links and population clustering which is commonly referred to as small world topology. Comparison with a hydrodynamic based network indicates that the genetic population network topology is similar. This approach shows the genetic structure of the *S. hystrix* coral follows a small world pattern which supports the results derived from previous hydrodynamic modelling.

**Keywords** coral dispersal; small world topology; networks; *Seriatopora hystrix*.

**1 Introduction**

The exchange of genetic information between reefs, through coral larvae exchange, is fundamentally important for the function of the Great Barrier Reef (GBR) (Hughes et al., 1999; Ayre and Hughes, 2004). Recruitment rates and recovery from disturbances are influenced by the migration of coral larvae that either settle locally or are transported across large distances (van Oppen and Gates, 2006). However the marine environment is not always favourable for coral growth and coral reefs are subject to natural disturbances which alter the community composition (Wakeford et al., 2008). In addition to natural disturbances, the rates and severity of anthropogenic disturbances, such as climate change, overfishing and coastal development (Hughes et al., 2003), are continuing to increase (Gardner et al., 2005; De'ath et al., 2009). In theory, high gene flow through larval connectivity could promote resilience and recovery in coral following disturbances (Hughes et al., 2003; van...
Oppen and Gates, 2006; Game et al., 2008; van Oppen et al., 2011). Measuring and understanding this complex network of gene flow is critical for conservation efforts (Cowen et al., 2002).

Due to the difficulties in measuring larval exchange by direct means (Palumbi, 2003), connectivity research is heavily reliant on hydrodynamic simulations (James et al., 2002; Werner et al., 2007). While the models have become more sophisticated in recent years (Legrand et al., 2006; Luick et al., 2007) the simulations of larval movements are limited by issues of spatial scale and how individual larvae behave (Baums et al., 2006). Critically, the interest in population connectivity is not limited to just the transport of larvae but also includes settlement success and juvenile mortality (Possingham and Roughgarden, 1990; Mumby, 1999; Bongaerts et al., 2010; Marshall et al., 2010). Hydrodynamic models therefore only reflect part of the recruitment process.

An alternative method of measuring the dispersal between distant reefs is to examine the genetic composition of coral populations in a number of localities in a region and then use statistical analysis of relationships to infer connectivity across that region (Hellberg, 2007; Underwood et al., 2007; van Oppen et al., 2008). DNA markers, like microsatellites, have sufficient variability to analyse how individuals relate to the sampled populations (Pritchard et al., 2000) yet the allele proportions at most loci are not distorted by local selection pressures. This has the advantage of informing us about the ultimate recruitment success rather than just larval movement.

Calculating the genetic distance between populations provides a mechanism to measure the likelihood of interbreeding between migrants and established individuals. The $F_{ST}$ metric commonly used to describe genetic differentiation implies that when two populations are not significantly different (i.e. $F_{ST} = 0.00$) that panmixia has occurred recently or is ongoing. Another measure, denoted by $D_{LR}$, is the mean genotype log-likelihood ratio across individuals from the two populations (Paetkau et al., 1997). $D_{LR}$ was found to differentiate fine scale population structure with high levels of confidence in empirical studies by Paetkau et al. (1997) and Underwood (2009).

This study describes the network created using the genetic analysis of the brooding pocilloporid coral *Seriatopora hystrix* Dana 1846. *S. hystrix* is a common coral in many parts of the world (Veron, 2000) despite being a predominantly self recruiting species. The life history of *S. hystrix*, in which eggs are fertilised internally and then released at an advanced developmental stage, ensures that larvae can settle within hours to days and this limits long distance dispersal (Ayre and Dufty, 1994; Underwood et al., 2007). The majority of larvae settle within 100 metres of the natal colony but a small percentage (2-6%) are successful in moving large distances (Underwood et al., 2007). The larvae are equipped with maternal zooxanthellae and could exist for many weeks in the water column (Underwood et al., 2007). Given the velocity of the currents within the GBR (James et al., 2002; Luick et al., 2007), settlement at distant reefs, forming a region-wide dispersal network, is conceivable.

The analysis of this network can utilise graph theory concepts and methods (Steuer and Lopez, 2008; Fortuna et al., 2009). Graph theory is experiencing a resurgence due to its successful application in complex systems (Green and Sadedin, 2005) including ecology (Urban and Keitt, 2001; Bascompte, 2009; Dormann, 2011; Ferrarini, 2011; Zhang, 2011a, b). The results of population differentiation analysis can be used to construct a network of vertices and lines. A connecting line can be formed from the centroid of the polygon containing the possible natal population to the centroid of the polygon containing the population where the migrant was found. The lines and centroids can then be used to construct a network. Kininmonth et al. (2010) have discovered that, based on hydrodynamic models, the Great Barrier Reef (central section) has a topology that can be described as “small world”. The question remains – does this same result hold true at a regional scale, for a particular species, where the connections are determined using genetic information.
Graphs that have a high level of clustering (where vertices are densely interconnected to form groups) and small average minimum path lengths (describing the number of links from one vertex to every other) compared with random or regular lattices are known as ‘small world’ graphs (Watts and Strogatz, 1998). Small world topology increases the connectivity of vertices, effectively making the entire network smaller, in terms of interactions, than a random or regular one. Clusters connected by longer links enable any point in the network to be located a small number of links from any other point and this forms the basis of Stanley Milgram’s six degrees of freedom (Milgram, 1967) for the network describing human social interactions (Barabasi, 2002).

We derive the $S. \text{hystrix}$ network and compare it to nine archetypal networks (Table 1). We then answer the following questions: What is the topology of the genetic distance network? What is the cluster structure within this topology?

2 Methods
In this section the construction and analysis of the network created by assigning individuals to a collection of surveyed populations of $Seriatopora \text{hystrix}$ will be described. We begin by describing the raw data and specimen collection process.

Van Oppen et al (2008) collected fragments from 1025 $S. \text{hystrix}$ colonies from 22 sites across the Great Barrier Reef and Coral Sea between March 2003 and February 2005. This sample set was later extended, during the years 2005 to 2006, to a total of 47 sites containing 2163 colonies (Fig. 1) (M. van Oppen, unpublished data). From each of these samples the genotypes at ten microsatellites were determined. The 36 to 54 samples collected at each site were contained within a 100 metre$^2$ area, consistent with the dispersal studies of $S. \text{hystrix}$ (Underwood et al., 2007).

The microsatellite composition formed the basis for quantifying the relationship among sites. Following the methodology described by Underwood (2009) the genotype likelihood ratio distance ($\text{Paetkau et al., 1997}$), $D_{LR}$, was calculated using the online calculator Doh (Brzustowski, 2002). For each combination of paired populations a $D_{LR}$ value was calculated. $S. \text{hystrix}$ is a common species known to occur across the region (Veron, 1986) and therefore many populations were not sampled. Hence the $D_{LR}$ value is a measure of how two populations are related irrespective of the number of intermediate populations that may have contributed to the connectivity.

The $D_{LR}$ values were transformed into graphs using tools created with the igraph package (Csardi and Nepusz, 2006) within the R environment (http://cran.r-project.org). The probabilities provided by the genotype likelihood ratio distance ($D_{LR}$) are taken to indicate the strength of connection. The line then has the attributes of connection strength, physical distance but not the direction of flow.

We addressed the initial question, is the genetic exchange of $Seriatopora \text{hystrix}$ within the GBR a small-world network, in the following manner. The values of average minimum path length ($L$), average degree (the average number of connections between vertices, $<k>$) and cluster coefficient (CI) were analysed for the network created by the individual exchange of $S. \text{hystrix}$ and compared with nine kinds of graph all of the same size (47 vertices), each created using a different process (Table 1). While the nine comparison networks have defined algorithms for construction it should be noted that the parameter selection will determine the topological outcome. For this comparison every network had a size of 47 vertices and the linkages were undirected.

The question of the clustering structure is addressed by examining the cluster membership of the $S. \text{hystrix}$ network. We utilise the leading eigenvector algorithm by Newman (Newman, 2006) to derive the clusters that
form the basis for the small world topology. This algorithm implemented using the igraph library in R (Csardi and Nepusz, 2006) utilises a modularity function (Newman, 2006) to optimise the number and membership of clusters in unweighted networks. However since the S. hystrix network is fully connected with weighted edges we needed to discard the weaker edges until the network showed optimal modularity whilst remaining connected (see method in Kininmonth, van Oppen et al. (2010)). The membership of each cluster can then be described.

Fig. 1 Map of the genetic survey sites (solid blue circles and labelled where feasible).
Table 1 Networks used for comparison. Vertex number is denoted by N.

<table>
<thead>
<tr>
<th>Network</th>
<th>Construction method</th>
<th>Parameters for Null Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric</td>
<td>First a number of points are dropped on to a unit square, then two points will be connected with an undirected edge if they are closer to each other in Euclidean distance than a given radius (Csardi and Nepusz, 2006).</td>
<td>N=47, radius =0.24, torus=TRUE</td>
</tr>
<tr>
<td>Forest Fire</td>
<td>This network model resembles how a forest fire spreads by igniting (connecting) trees close by Leskovec et al. (2007). Vertices are added sequentially and edges are created with respect to the neighbouring configuration.</td>
<td>N=47, forward probability=0.39, backward burning ratio=0.32/0.37</td>
</tr>
<tr>
<td>Watts-Strogatz</td>
<td>First a lattice is created with the given dimension, final size and neighbourhood size arguments. Then the edges of the lattice are rewired uniformly randomly with set probability (Watts and Strogatz, 1998).</td>
<td>Dimension of starting lattice =1, N=47, neighbourhood =5, rewiring probability =0.00,0.02, 0.05, 0.1</td>
</tr>
<tr>
<td>Barabasi-Albert scale free model</td>
<td>One vertex is added in each time step and edges are then created to link existing vertices with preferential attachment, ( P(k) \sim k^{-\gamma} ) (Barabasi and Albert, 1999) where ( k ) is the degree and ( \gamma ) is a constant.</td>
<td>N=47, preferential attachment power = 1, number of edges to add in each time step = ( \sqrt{47} )</td>
</tr>
<tr>
<td>Star</td>
<td>This algorithm creates a star graph, in which every single vertex is connected to the centre vertex and no other (Csardi and Nepusz, 2006).</td>
<td>N=47</td>
</tr>
<tr>
<td>Lattice</td>
<td>This pattern is represented by a lattice of 2 dimensions composed of a replicating set of four vertices joined in a square (Csardi and Nepusz, 2006).</td>
<td>Dimensions of lattice = ( \sqrt{47} ) by ( \sqrt{47} )</td>
</tr>
<tr>
<td>Erdos-Renyi Random</td>
<td>Initially the vertices are created then the set number of edge are allocated randomly (Erdos and Renyi 1959).</td>
<td>N=47, number of edges=230</td>
</tr>
<tr>
<td>Tree</td>
<td>This model is based on a regular tree with 3 edges connecting the ‘child’ neighbours (Csardi and Nepusz, 2006).</td>
<td>N=47, children = 3</td>
</tr>
<tr>
<td>Complete</td>
<td>This graph has every vertex connected to every other vertex (Csardi and Nepusz, 2006).</td>
<td>N=47, self loops = FALSE</td>
</tr>
</tbody>
</table>

3 Results
What is the topology of the genetic exchange of *S. hystrix* network? The genetic survey sites (Fig. 1) using the \( D_{LR} \) values were used to create a fully connected unidirectional network containing 47 vertices and 1081 edges (Fig. 2). The edge weight distribution is also shown in Fig. 2. The weight value is based on the genetic distance and so small \( D_{LR} \) values indicate that the populations are genetically similar. The histogram shows that only a small number of edges (<130) has low \( D_{LR} \) values (<5).

In order to investigate the dominant topology we removed the weaker edges (higher \( D_{LR} \) values) sequentially and recorded the cluster coefficient and the average minimum path. The value of these metrics changes as the weaker edges are removed (Fig. 3). Comparing to a selection of null models (Table 1), Fig. 3 shows the topology shifts towards the small world models constructed by the Watts-Strogatz and Geometric algorithms.

To investigate the community structure of the genetic network we used the optimum modularity value (Fig. 4) to select the threshold that best defines the network modules. This point can be observed in Fig. 4, where only 307 edges remained with weights less than 9.1 (Fig. 5). There was one connected component (connected section of network) and two isolated populations (Osprey Reef and Ribbon Reef 53). Removing edges below
this threshold quickly fragments the network (Fig. 4). Based on this edge threshold of 9.1 (modularity = 0.45 for 3 components) the network has a cluster coefficient of 0.76, an average minimum path length of 2.2 and an average degree of 13.0.

Fig. 2 Histogram of the edge weight distribution (a) and network of the DLR linkages showing the fully connected structure (b).
Fig. 3 Plot of how the genetic network of *S. hystrix* (open circles) changes as weaker connections are removed, and how it compares with the nine network models (Table1 – closed circles). Edges with weights greater than the threshold (selected thresholds = 1.0 (dlr 1.0), 1.4 (dlr 1.4), 2.2 (dlr 2.2), 4.7 (dlr 4.7), 9.1 (dlr 9.1), 14.8 (dlr 14.8)) were removed. The fully connected network is in the far right (same as 'complete' model) and the network with most edges removed (dir 1.0) is in the far left.
Fig. 4 Plot of the modularity per component as the edge weight threshold is changed from 40 to 0, effectively removing the weaker edges (note: high edge weight=high $D_{LR}$ value=weakly associated populations). The lines indicate an optimal modularity of 0.45 (components = 3) at a threshold edge weight of 9.1.
Given these graph theory metrics, let us first consider how this network compares to the hydrodynamic small-world networks of the GBR (Kininmonth et al., 2010). Kininmonth et al. (2010) showed that the larval dispersal network of the GBR over the 30 year period has an average minimum path length between 2.2 and 3.9 with an average degree ranging from 67 to 128 (noting that the number of vertices was 321 so the average density of connection was 0.32). This makes the GBR hydrodynamic network similar to the genetic network presented here (average degree of 13 with 45 vertices equals a density of 0.28, average minimum path length = 2.2). However the cluster coefficient values differ with the hydrodynamic network ranging from 0.26 to 0.31 while the genetic network was 0.76. This indicates that the genetic network contained relatively more triangular loops than the hydrodynamic network (adjusted for directional links) implying higher clustering at a local scale.
We compared the nine null models (Table 1, Fig. 3) to the *S. hystrix* network (at 9.1 or less edge weight). The *S. hystrix* network was more clustered but with a similar average minimum path length than the small world models based on the Geometric and Watts Strogatz algorithms. The *S. hystrix* network has a smaller average minimum path length than the Forest Fire model but larger than the Erdos Renyi and Barabasi Albert models. This highlights the relative proportion of long distance connections that serve to reduce the path length. The Tree, Lattice and Star models have no triangular clusters but the *S. hystrix* network has an average minimum path length half way between the Star and Lattice models. This implies that the *S. hystrix* network topology is not based on a simple diffusion model of propagule dispersal but also contains longer distance dispersal from currents.

The number of clusters for the *S. hystrix* network was found to be 3 (Fig. 5) when optimised for the modularity and based on the eigenvector algorithm. The relationship of the cluster membership to the geographic position is not clearly defined but there are broadly northern, central and southern groups. In particular for the “green” cluster some central reefs (Myrmidon, Davies and Broadhurst) are associated with the northern outer reefs such as the Ribbon reefs. The “red” coloured cluster includes reefs that extend from Heron Island to Lizard Island. The “blue” cluster is exclusively composed of southern reefs including those at the extreme latitude of 30 degrees south.

3 Discussion
We have discovered that the *S. hystrix* larval exchange network on the eastern coast of Australia, as determined by genetic relatedness, exhibits a small world pattern. This is consistent with our analysis of the connectivity properties of the GBR which did not relate to any specific species and inferred connectivity using hydrodynamic data only (Kininmonth et al., 2010). The implication from this finding is that, despite the highly localised recruitment strategy of *S. hystrix* (Underwood et al., 2009), the distant populations remain connected through rare long distance dispersal. Thus the entire *S. hystrix* metapopulation in the study region is more connected through a small world topology than the dominating local recruitment strategy would suggest.

The small world topology implies that the *S. hystrix* populations are resilient to local and mesoscale disturbances (Watts and Strogatz, 1998; Gaedke, 2008). In particular the recovery from severe disturbance for any particular site will be largely determined by the site’s location in the larval dispersal network. Underwood et al. (2009) found that limited recruitment dramatically reduced the recovery of reefs following severe disturbance. If a region is severely impacted by bleaching, like the 2001/2 event (Berkelmans et al., 2004), then migrating larvae will need to arrive via the long distance water currents rather than the nearby reefs. With an average minimum pathway length of 2.2, larvae from distant reefs in the GBR (with potentially a more robust suite of alleles (van Oppen and Gates, 2006)) could arrive in less than 3 migration events. This property of being closely connected to distant vertices through intermediate clusters is typical of small-world networks. At a larger scale, Treml et al. (2008) used an Eulerian advection–diffusion approach to measure connectivity across the Pacific reefs and found that the population clusters could be weakly connected by long distance migration. Our research supports this perspective that coral and fish can be connected as a mesoscale metapopulation.

The *S. hystrix* genetic network could be described as three clusters but what does this really imply? Allocating each sample population into an artificially defined grouping based solely on the microsatellite composition of the individual colonies has the power to describe the character of the metapopulation dynamics (Fortuna et al., 2009). For a brooding coral, such as *S. hystrix*, the strongly defined clusters highlight the recruitment processes being predominantly local (Underwood et al., 2009) but influenced with occasional long distance events. These clusters, as defined in Fig. 5, have a higher density of connections within the cluster than between the clusters. In many respects this relationship is a direct function of the physical location of the
reefs within a complex hydrodynamic matrix. The far southern reefs such as Middleton, Elizabeth and Lord Howe (Fig. 1) rely on strong currents to sweep the larvae across approximately 1000 kilometres of open ocean (Noreen et al., 2009). The linkages between the southern cluster and the two GBR clusters highlight the fragile and infrequent bonds within the metapopulation especially in comparison to the bonds between the northern and central clusters. However this cluster model challenges the latitudinal and cross-shelf orientated divisions that presently group the GBR (Kerrigan et al., 2010). Instead the dense matrix of coral reefs, that give support to the name Great Barrier Reef, provides a highly interactive environment for complex metapopulation dynamics.

Determining where to sample, in a project such as this, requires a particularly difficult balance of resources and scientific focus. Sampling many populations in a small region will provide insight into the clustering character for *S. hystrix* but will be ineffective in understanding the long distance dispersal events. Ayre and Dufty (1994) found in their genetic sampling of *S. hystrix* in the GBR that each reef forms discrete populations that are infrequently linked by migration. The lagoonal sites, with minimal access to strong currents, tend to be genetically different to those in exposed locations (Ayre and Dufty, 1994; Underwood et al., 2007; van Oppen et al., 2008). To complicate matters further, sampling at multiple depths can also reveal strong genetic structure (Bongaerts et al., 2010; van Oppen et al., 2011). At a regional scale, sampling the populations that are isolated from the main reef complex, such as Osprey Reef (Fig. 1), highlight significant genetic distinctiveness (van Oppen et al., 2008; Noreen et al., 2009). Sampling within the strongly connected clusters such as the mid-shelf reefs (i.e. Emily, Tongue and Sudbury Reefs), focuses on populations with strong multiple connections to populations in the region. Stratified sampling, informed by the dispersal network (inferred from hydrodynamic modelling), will reveal the metapopulation structure.

The characterisation of the network as a larvae exchange system depends on more than just the topology but also includes the rate of exchange between populations. The strength of connection is a significant parameter used to model metapopulation dynamics (Roughgarden et al., 1988; Armsworth, 2002). However directly gauging the strength of larval exchange through field based surveys is a particularly difficult task (Palumbi, 2003; Jones et al., 2005). Empirical data are often lacking for the survival rate of each development stage for the migrating individuals and estimation of larval transport is only a small component of the metapopulation dynamics (Armsworth, 2002).

The power of a genetics study, such as this one, is that it provides information about real historical connections as determined by the genetic information in individuals. In contrast, measuring the water currents with a variety of instruments, and then translating this to Lagrangian particle movements, requires the acceptance of many assumptions known to be false or misleading. For instance in the GBR reef connectivity study by James et al (2002) larvae were released uniformly as a function of reef edge. The GBR reefs are complex structures with structural features resulting from wave and current influences (Hopley, 1982) and species, like *S. hystrix*, will not be uniformly distributed (Veron, 1986).

While genetic studies, require considerable field work and expensive analytical capacity the empirical data derived can directly target the key questions relating to metapopulation dynamics (Hellberg, 2007). How strongly connected are sites at a variety of scales? What influence does migration have on recovery after disturbance? Where should conservation efforts be focused in order to enhance the resilience of the coral and fish metapopulations?

The apparent high level of connectivity raises the question that if the *S. hystrix* larval exchange is a well connected small world network then why is there such high genetic differentiation between populations? Despite the exploration of the mesoscale genetic structure of coral being in its infancy we present the following hypothesis: local scale disturbances regularly reduce the relatively homogenous self-seeded
populations (in both colony number and fecundity (Hughes et al., 2000; Wakeford et al., 2008), resulting in an increase in the effective recruitment rate of migrant larvae where the rate is a function of the neighbourhood dispersal topology. Importantly, the migration rate is a function of the inclusion within the heterogeneous larval exchange network.

This analysis has implications for coral reef system management. Conservation of regional biodiversity will be enhanced by the identification and protection of the reefs that maintain the established network (Almany et al., 2009). Protection measures, like marine protected area networks, will be more effective if focus is directed to reefs that provide source larvae, act as stepping stones linking distant clusters or are isolated (Palumbi, 2003; Kininmonth et al., 2011). Increasingly the emphasis for conservation is on the establishment of long term measures to ensure the persistence of populations in the face of catastrophes (Game et al., 2008; Game et al., 2008) and anthropogenic pressures (Lubchenco et al., 2003). Only by understanding the role of each population within the framework of dispersal related resilience can conservation measures hope to be successful (Hodgson et al., 2009; Beger et al., 2010).

Acknowledgements
Particular thanks are due to Jim Underwood for the DLR calculations and the rest of the genetic’s team at the Australian Institute of Marine Science for the provision of the S.hystrix genetic data. Special acknowledgement is due to Tamas Nepusz and Gabor Csardi for the igraph package used for network analysis.

References
116-119


