RESULTS OF AN EXPLORATORY STUDY OF THE APPLICATION OF AGENT-BASED MODELLING OF SHEEP GENETIC PROGRESS USING NETWORK THEORY

A. Reverter[1], S. Dominik[2], S. Field[3] and R. Banks[4]

[1]CSIRO Livestock Industries, 306 Carmody Rd., St. Lucia, Brisbane, QLD 4067, Australia
[2]CSIRO Livestock Industries, McMaster Laboratory, Chiswick, Armidale, NSW 2350, Australia
[3]Sheep Genetics, PO Box U254, University of New England, Armidale, NSW 2351, Australia
[4]Meat & Livestock Australia, C/- University of New England, Armidale, NSW 2351, Australia

SUMMARY

With the overarching aim of better managing genetic progress at an industry level, this study explored whether coupling the power of agent-based models with the elegance of network interaction topology can assist with the two-fold aims of: (1) characterising the relationships existing in the Australian sheep breeding industries; and (2) improving the development and delivery of decision aids and tools for sheep breeders. Data from the August 2010 LAMPLAN evaluation was interrogated. Input and output files from the genetic evaluation of Poll Dorset and White Suffolk were processed to generate a network where nodes were flocks and edges connecting nodes represented the sharing of genetic material via common sires. As a result, we report on the interplay between a series of flock attributes including size, sex mating ratio and network connectivity structure with CarcasePlus Index value.

INTRODUCTION

Agent-based modelling aims at using decision-making rules to model the actions and interactions of autonomous agents, both individual (eg. at the flock level in our context) or collective (eg. at the breed level), with a view to re-create and predict the appearance of complex phenomena. It combines elements of game theory, complex systems, computational biology, and evolutionary programming. Bonabeau (2002) provides an introduction of the basic principles of agent-based models and argues that its real-work application can be encapsulated in four main areas: flow simulation, organisational simulation, market simulation and diffusion simulation.

Network theory exploits interactions in terms of nodes and edges. In our context, nodes could be flocks (commercial and stud), and edges could be the relationships between them, eg. via the sharing of genetic material. Within the context of molecular biology, Barabási and Oltvai (2004) presented a landmark review outlining the most basic network architectural measures including degree distribution, clustering coefficient and path length. These three measures alone allow distinguishing random from non-random networks.

The aim of this paper was to conduct an initial examination of the value of coupling agent-based models with network theory to better characterise genetic progress.

MATERIALS AND METHODS

Data and edits. Data from the August 2010 LAMBPLAN evaluation was downloaded from the Sheep Genetics database (Sheep Genetics 2011). Input and output files from the OVIS analysis (Brown *et al.* 2001) corresponding to Poll Dorset and White Suffolk were processed. The initial dataset, comprising >1.6M animals from ~1,300 flocks, was edited to include only records from fully pedigreed individuals with date of birth available and from flocks with 11 consecutive years of records from 1999 to 2009. For the 492,776 sheep (280,950 Polled Dorset and 211,826 White Suffolk) in 145 flocks (73 Polled Dorset and 72 White Suffolk) fulfilling these editing criteria, OVIS results corresponding to $index8 (the "CarcasePlus" index) were retrieved. Among the 145 flocks there were 38, 48, 7, 37 and 15, from NSW, SA, TAS, VIC and WA, respectively.

Flock attributes. For every flock in the edited dataset, we defined seven attributes as follows: (1) Size = Number of animals registered; (2) MatRat = Average mating ratio (ie. Females per male); (3) TotConn = Total number of connecting flocks (ie. Flocks with whom sires are being shared); (4) HiConn = Connections to flocks with higher average 2010 carcase EBV; (5) LoConn = Connections to flocks with lower average 2010 carcase EBV; (6) CarcEBV = Average 2010 CarcasePlus index and (7) ProgEBV = Average genetic progress based on the regression of CarcasePlus index on year of birth for years 1999 to 2009. The last two attributes were used as indicators of flock genetic performance. Also, for every pair of connecting flocks we recorded the number of sires in common over the same 11 year period. This set of attributes was chosen simply to allow exploration of the data: it is by no means the definitive set of all attributes of a network and its components that could be examined.

Network construction and visualisation. Flock-to-flock interactions were processed to generate a network where nodes were flocks and edges connecting nodes represented the sharing of genetic material via common sires. To visualise the resulting network, we used the Cytoscape software (Shannon *et al*. 2003; http://www.cytoscape.org) where the above-mentioned attributes were also incorporated in the visualisation schema.

RESULTS AND DISCUSSION

Table 1 shows summary statistics for the flock attributes. The two indicators of performance (2010 average CarcasePlus index and genetic progress over the period 1999 to 2009) were moderately correlated with each other ($r = 0.474$; $P < 0.001$). This moderate correlation persisted when the two breeds were considered separately (Figure 1A) indicating the two indicators are complementary measures of performance. Also, larger flocks were associated with higher average index value ($r = 0.451$; $P < 0.001$). However, this relationship vanished when the actual genetic progress was used as indicator of performance ($r = 0.072$; $P > 0.05$). Similarly, there was a positive association between degree of connectedness and genetic merit: highly connected flocks had higher genetic performance regardless of the indicator used, while less connected flocks tended to have lower average CarcasePlus index. The separation of the HiConn and LoConn suggests that the performance of the partners in a connection is of importance. Flocks with lots of connections to low EBV flocks appear to have higher genetic performance and vice versa. This result could be attributed to having ignored the flock of origin of the sire(s) involved in the connections. In simple terms, high performing flocks are acting as "*donors*" to many lower performing flocks.

Table 1. Summary statistics for the attributes of the 145 flocks included in this

Attribute[A]	Summary Statistics				Correlation with[B]	
	Mean	SD	Min.	Max.	CarcEBV	ProgEBV
Size	3,398	2,060	184	14,393	0.451	0.072[NS]
MatRat	16.43	7.72	2.10	41.87	-0.302	-0.229
TotConn	57.03	24.96	1.00	119.00	0.588	0.394
HiConn	28.52	17.84	0.00	82.00	-0.396	-0.077[NS]
LoConn	28.517	25.154	0	93	0.864	0.446
CarcEBV	148.793	8.282	130.810	173.967	1.000	0.474
ProgEBV	4.658	1.127	1.625	8.385	0.474	1.000

[A]See Materials and Methods for definition of flock attributes.
[B]Correlation values with an "NS" superscript are not significantly different from zero ($P > 0.05$).

For the flock network studied, Figure 1B shows the power-law scale-free distribution of the number of connections as a function of the number of sires represented in a connection. The vast majority of connections are represented by a single or a few sires, while very few connections are represented by lots of sires. Also, all 145 flocks were connected to at least one other flock and on average they were connected to 57.03 flocks, with a range from 1 to 119 flocks (Figure 1C).
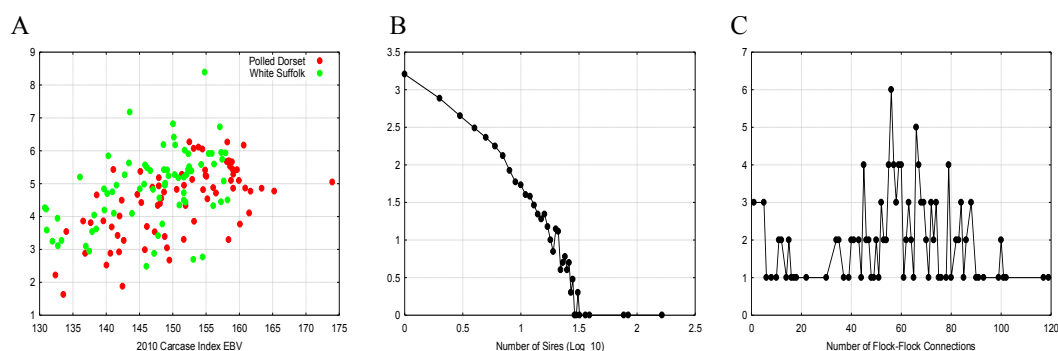
Figure 1. A: Relationship between 2010 carcase index EBV and genetic progress from year of birth 1999 to 2009 and by each breed (red = Polled Dorset; green = White Suffolk); B: Frequency of flock to flock connections by number of sires represented in each connection; C: Frequency of flocks as a function of the number of connections.

The network generated with the 145 flocks contained 4,135 edges. This represents a clustering coefficient of 39.61% indicating the percentage of the total number of possible connections that could exist with 145 nodes (ie 100% would mean all flocks were connected to all flocks).

After imposing a filtering criterion to only include those flock-to-flock connections represented by at least 10 sires, we obtained the visualisation schema presented in Figure 2 with 83 flocks and 322 connections (ie. Clustering coefficient = 9.46%). The visualisation schema shows a clear separation between the two breeds (red = Polled Dorset; green = White Suffolk). At the kernel of the network we reveal a White Suffolk flock from WA (flock ID = 23_0090_WA) of medium size (6,913 animals in the dataset) and rapid genetic progress (5.91 index units per year). This flock provides a key pathway between the two populations (breeds).

CONCLUSIONS

The network analysis approach provides a useful tool to visualise the characteristics of individual flocks and the relationships between flocks, both defined through a number of parameters. It highlights key flocks that connect large parts of the industry. As indicated by the width of their nodes outline, highly connected flock have mostly higher genetic progress, but also flocks can be identified that perform worse than other connected flocks. This would likely indicate that these flocks are not making optimal selection decisions when sourcing and/or selling sires.

The present study represents a first attempt to explore the attributes that should be considered when the intention is to perform agent-based modelling in a network theory framework and applied to genetic progress in sheep breeding systems. This work tackles an important problem: understanding the Australia-wide sheep genetic improvement system, and informing future breeding / management decisions using state-of-the-art methods. Further work is required to fully exploit the power is the proposed approach in particular with respect to the identification, measurement and simulation of the attributes within the context of agent-based models.
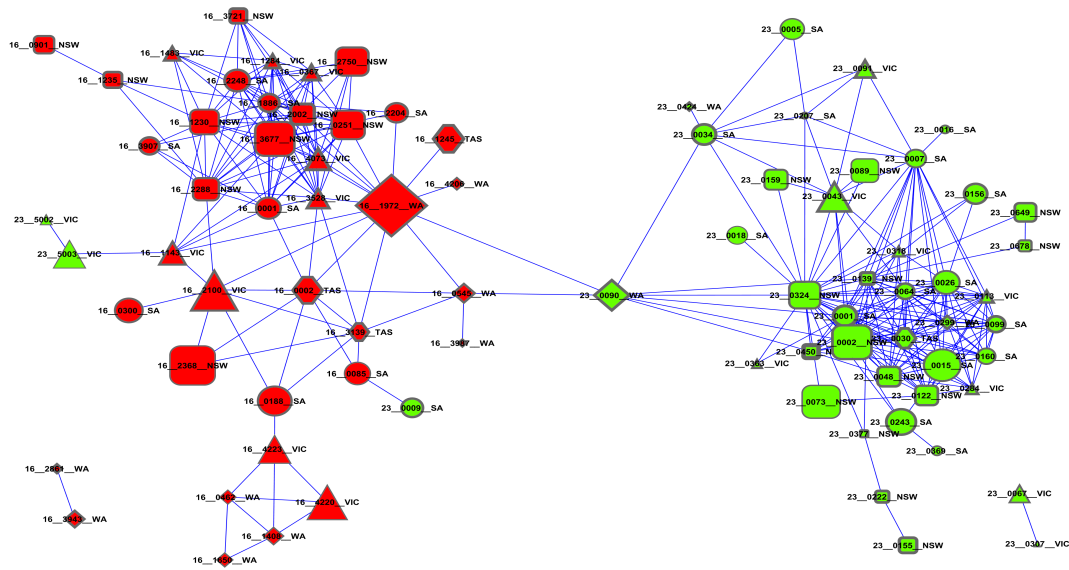
Figure 2. Network generated by 83 flocks and 322 connections where connections are represented by at least 10 sires. Red and green nodes represent Polled Dorset and White Suffolk flocks, respectively. Node size represents flock size. Node shape represent origin with NSW, SA, TAS, VIC and WA represented by rectangles, ellipses, hexagons, triangles and diamonds, respectively. Finally, node line width indicates genetic progress.

In spite of these limitations, three critical outcomes were identified: (1) The tendency for highly connected flocks to achieve higher genetic performance; (2) The emergence of 'hub' flocks providing inter-breed relationships; and (3) The identification of problematic flocks. Importantly, we never asked the question "Which, if any, inter-breed flock is highly connected to other flocks and yet has an average size but large genetic progress?" Instead, this information emerged as a natural phenomenon of the approach undertaken. This type of "naturally emerging" information can be used to manage genetic progress better at an individual and at an industry level.

We conclude that network analysis may help individuals and organisations involved in sheep (and other species) genetic improvement, understand and think about the system in new ways, and on this basis, the approach warrants further investigation.

REFERENCES
Bonabeau E. 2002. *Proc. Natl. Acad. Sci*. **99**:7280.
Barabási A.-L. and Oltvai Z.N. 2004. *Nat. Rev. Genet*. **5**:101.
Brown D.J., Tier B., Reverter A., Banks R. and Graser H.-U. 2001. *Wool Tech. Sheep Breed*. **48**:285.
Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B. and Ideker T. 2003. *Genome Res*. **13**:2498.
Sheep Genetics (2011) Sheep Genetics, from http://www.sheepgenetics.org.au, Date Accessed: 15/4/2011