Vol 455 | 23 October 2008 | doi:10.1038/nature07366

LETTERS

Galaxies appear simpler than expected

M. J. Disney[1], J. D. Romano[1,2], D. A. Garcia–Appadoo[3,1], A. A. West[4,5], J. J. Dalcanton[5] & L. Cortese[1]

Galaxies are complex systems the evolution of which apparently results from the interplay of dynamics, star formation, chemical enrichment and feedback from supernova explosions and supermassive black holes[1]. The hierarchical theory of galaxy formation holds that galaxies are assembled from smaller pieces, through numerous mergers of cold dark matter[2–4]. The properties of an individual galaxy should be controlled by six independent parameters including mass, angular momentum, baryon fraction, age and size, as well as by the accidents of its recent haphazard merger history. Here we report that a sample of galaxies that were first detected through their neutral hydrogen radio-frequency emission, and are thus free from optical selection effects[5], shows five independent correlations among six independent observables, despite having a wide range of properties. This implies that the structure of these galaxies must be controlled by a single parameter, although we cannot identify this parameter from our data set. Such a degree of organization appears to be at odds with hierarchical galaxy formation, a central tenet of the cold dark matter model in cosmology[6].

About 300 sources, from part of the much larger, blind 21-cm survey for neutral hydrogen made with the Parkes radio telescope[7–9], overlap a region surveyed by the Sloan Digital Sky Survey (SDSS) in the optical spectral region[10]. Two hundred were unambiguously identified as individual galaxies[11], and are representative of the whole range of galaxies between giant spirals and extreme dwarfs, missing only the ~10% of largely neutral-gas-free galaxies found mainly in big clusters. For each object, we measured the H I (neutral hydrogen) mass and line width ΔV, redshift, inclination, two radii (R_{50} and R_{90}) respectively containing 50 and 90% of the emitted light, the luminosity L_g and four colours. Three of the colours are degenerate[11,12], leaving dynamical mass ($M_d \equiv (\Delta V)^2 R_{90}/G$; ref. 13), H I mass ($M_{\rm H\,I}$), luminosity, radius, concentration (R_{50}/R_{90}) and colour (SDSS (g − r)). For comparison, we can imagine galaxies being controlled by seven physical quantities, namely total mass, baryon fraction, age, specific angular momentum, specific heat energy (random motion), radius and concentration, only six of which can be independent, owing to the Virial theorem. We thus have as many independent observables as we do controlling physical parameters, making an examination of the correlation structure potentially very interesting. Each discovered correlation, if independent of the rest, will set a further constraint on galaxy physics.

We find five correlations, four of which were already known: that between dynamical mass and luminosity ($M_d \propto L_g$)[14], that between luminosity and colour (luminous galaxies are redder)[15], that between R_{50} and R_{90} (H I galaxies have exponential profiles)[16], and that between $M_{\rm H\,I}$ and R_{50} (all the galaxies have the same H I surface density, $M_{\rm H\,I}/R_{50}^2$)[11,17,18]. The new correlation, namely that surface brightness $\Sigma \equiv L_g/R_{50}^2$ is proportional to R_{50} (ref. 11), required a blind H I survey to clearly separate it from the pronounced selection effects in the optical. All the data and the correlations appear in ref. 11.

Here we examine the correlation structure by means of a principal component analysis (PCA)[19,20] based on the correlation matrix of the measured data. Being normalized (unlike the covariance matrix), a correlation-based analysis is immune to the influence of scaling. PCA can be thought of as a search in the six-dimensional space of observables for a smaller number of coordinates that describe nearly all the variance. For instance, PCA has been used to show that elliptical galaxies lie on a 'fundamental plane'[21,22], that is, in a two-dimensional space. Because such a correlation analysis relies on linear relations between variables, we use logarithmic quantities (the colour, being a magnitude, is also logarithmic)[23].

Colour turns out to be more complex than the other observables, so we omit it at first and reintroduce it later. Figure 1 demonstrates the strong correlations that exist between the five other variables. This is emphasized in Fig. 2, where all are seen to be strongly correlated with the first principal component, PC1, and scattered with respect to the other principal components. A high degree of organization is already evident. The eigenvalue of PC1 is 4.1, in comparison with a maximum

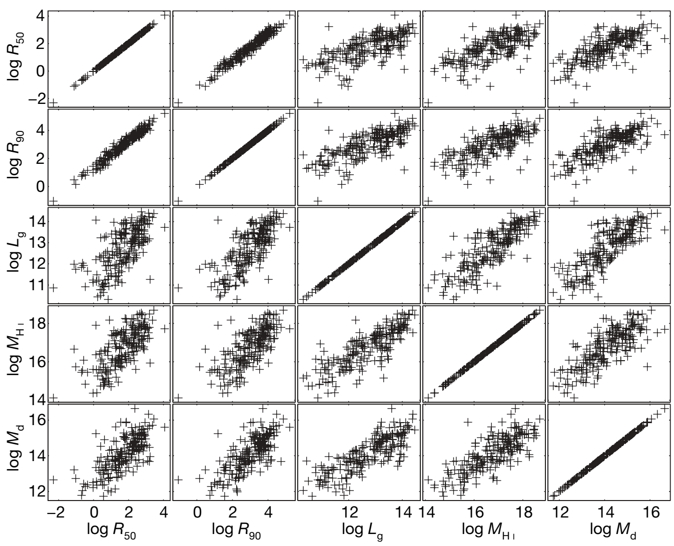

Figure 1 | Scatter plots showing correlations between five measured variables, not including colour. The variables are two optical radii, R_{50} and R_{90} (in parsecs), respectively containing 50 and 90% of the emitted light; and luminosity, L_g; neutral hydrogen mass, $M_{\rm H\,I}$; and dynamical mass, M_d (inferred from the 21-cm linewidth, the radius and the inclination in the usual way; see main text), all in solar units. In this normalized, logarithmic representation, the slopes are irrelevant; only the scatters are significant. We find that (see ref. 11) $R_{90} \propto R_{50}$, $L_g \propto R_{50}^3$, $M_{\rm H\,I} \propto R_{50}^2$ and $M_d \propto L_g$. Alternatively, the surface brightness $\Sigma \equiv L_g/R_{50}^2 \propto L_g^{1/3} \propto R_{50}$, whereas the luminosity density (defined as L_g/R_{50}^3) and mass density (defined as M_d/R_{50}^3) are independent of size. Colour is included in Fig. 3.

[1]School of Physics & Astronomy, Cardiff University, 5 The Parade, Cardiff CF24 3AA, UK. [2]Department of Physics and Astronomy, The University of Texas at Brownsville, 80 Fort Brown, Brownsville, Texas 78520, USA. [3]European Southern Observatory, Alonso de Cordova 3107, Casilla 19001, Vitacura, Santiago 19, Chile. [4]Astronomy Department, University of California, 601 Campbell Hall, Berkeley, California 94720-3411, USA. [5]Department of Astronomy, Physics-Astronomy Building C309, University of Washington, Box 351580, Seattle, Washington 98195, USA.

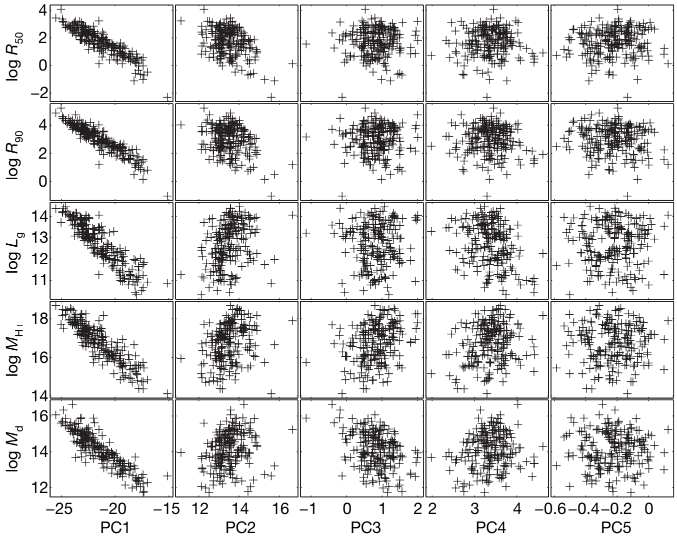

Figure 2 | Scatter plots showing correlations between the five measured variables and the principal components. The five physical quantities are tightly correlated with a single principal component, PC1, which accounts for their correlations with one another. The principal components are labelled from the strongest, PC1, to the weakest, PC5, with eigenvalues 4.1, 0.53, 0.23, 0.17 and 0.02, respectively. Eigenvalues less than 1 are normally thought to represent scatter only—but see main text. The direction cosines for PC1 relative to the physical variables in order plotted (top to bottom) are −0.45, −0.46, −0.44, −0.44 and −0.44 (that is, approximately $-1/\sqrt{5}$). This is expected for such a strong principal component and makes it impossible to single out any one observable as the 'driving force'. In any case, correlations alone should not be used to infer causation.

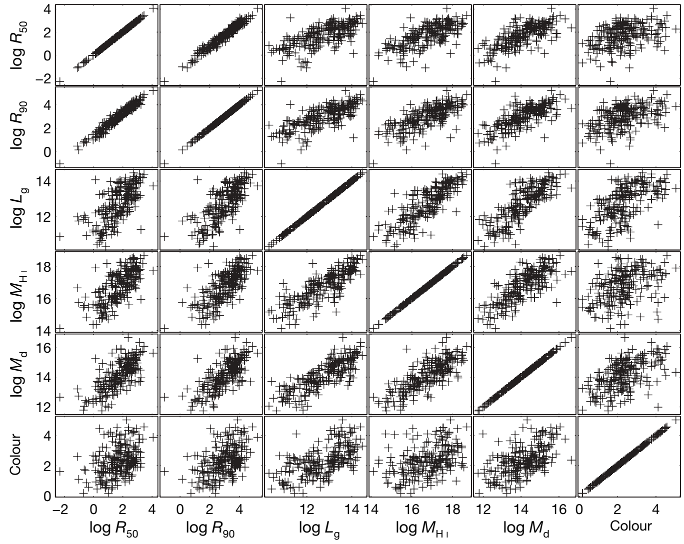

Figure 3 | Scatter plots showing correlations between the previous five measured variables (Fig. 1), but now including colour. The colour is (g − r), that is, 'green minus red', and is also logarithmic. Colour (measured within R_{90}) is correlated with the five other observables in the sense that more luminous galaxies are redder, which is usually taken to mean that their output is dominated by older stars. However, the colour correlations are much weaker, though still highly significant (at the 0.01% level).

possible of 5.0 (1 for each variable; they should sum to 5), whereas the next principal component has an eigenvalue of only 0.53. We note that all five correlations are independent because each introduces at least one a priori independent observable not present in any preceding correlation. We note further that each implies some new intrinsic (not merely scaling) property that calls for a physical explanation. For instance, the correlation between surface brightness and R_{50} implies that $L_g/R_{50}^2 \propto R_{50}$, or that the luminosity density (defined as L_g/R_{50}^3) is independent of either luminosity or dynamical mass.

Figure 3 shows the correlations when the colour is included. Colour is evidently correlated with all the remaining variables, but more weakly so. (The least significant such correlation (0.39) is nevertheless significant at the 0.01% level). Figure 4 clarifies the situation by exhibiting the principal component structure. Like all the other variables, colour is definitely correlated with PC1, which now has an eigenvalue of 4.4 out of a maximum possible of 6.0 (that is, explains 73% of the variance). The next highest eigenvalue is 0.75, for PC2 (Fig. 4, column two). Statisticians normally ignore, as insignificant, principal components with eigenvalues of less than 1.0, except in the special case in which one observable is scattered mostly independently of the rest, when values as low as 0.7 may be considered significant[19]. This turns out to be the situation here (0.75). There is a component of the colour that is correlated with nothing but itself, and this 'rogue' component constitutes a weak, but nevertheless significant, principal component, PC2, as is clear from column two. The strong correlation in the colour plot of column two is notable. Beyond the fact that PC2 is strongly aligned with colour (direction cosine, 0.83), what does this mean? Simulations (Supplementary Information) show that if colour were perfectly uncorrelated with everything else (that is, were randomly scattered), then the colour plot of column two would be a sharp, straight line at a slope of 45°. This is because even if an observable (for example colour) is correlated with nothing else it is nevertheless correlated perfectly with itself, and the PCA seeks out a principal component aligned perfectly with that observable (in this case colour).

We can summarize the PCA as follows. H I-selected galaxies appear to have colours made up of two components, which we label the 'systematic' component and the 'rogue' component. The rogue colour

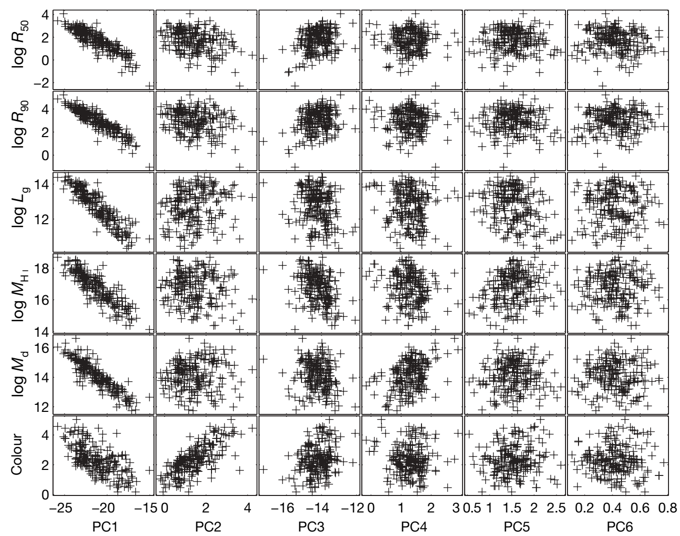

Figure 4 | Scatter plots showing correlations between all the six measured variables (including colour) and the principal components. The principal components are labelled from the strongest, PC1, to the weakest, PC6, with eigenvalues 4.4, 0.75, 0.48, 0.22, 0.14 and 0.02, respectively. The bottom plot of column one shows that colour is well correlated with the other five observables and with PC1. However, the bottom plot in column two shows that colour is even more strongly correlated with a new principal component, PC2, which is correlated with nothing else. If colour were entirely independent of all the other variables, it would have its own principal component with an eigenvalue of 1.0, instead of 0.75 as observed, and the data plotted against PC2 would lie on a sharp, straight line, because they would be correlated perfectly with themselves. The colour of a galaxy is evidently composed of two components: a 'systematic' component correlated with all the other observables, and a 'random' or 'rogue' component correlated with nothing but itself. Appreciation of this non-intuitive point was greatly assisted by examining a PCA of simulated data (see Supplementary Information).

is scattered more or less randomly, and is correlated with none of the other observables. (It would be natural, although not compelling, to identify the systematic colour with the old population of stars, and the rogue component with recent and transitory bursts of star formation—very luminous young stars being far bluer and shorter lived than older ones.) The remaining six variables, including the systematic colour, are all correlated with one another and with a single principal component. In other words, they form a one-parameter set lying on a single fundamental line. Such simplicity was unpredicted and is noteworthy when galaxies might have been controlled by any six of the seven potentially independent physical parameters listed earlier. It is even more significant considering the enormous variety among the galaxies in the sample[11].

If, as we have argued, galaxies come from at most a six-parameter set, then for gaseous galaxies to appear as a one-parameter set, as observed here, the theory of galaxy formation and evolution must supply five independent constraint equations to constrain the observations. This is such a stringent set of requirements that it is hard to imagine any theory, apart from the correct one, fulfilling them all. For instance, consider hierarchical galaxy formation in the dark matter model, which has been widely discussed in the literature[3,4]. Even after extensive simplification, it still contains four parameters per galaxy: mass, spin, halo-concentration index and epoch of formation. Consider spin alone, which is thought to be the result of early tidal torquing. Simulations produce spins, independent of mass, with a log-normal distribution. Higher-spin discs naturally cannot contract as far; thus, to a much greater extent than for low-spin discs, their dynamics is controlled by their dark haloes, so it is unexpected to see the nearly constant dynamical-mass/luminosity ratio that we and others[14] actually observe. Hierarchical galaxy formation simply does not fit the constraints set by the correlation structure in the Equatorial Survey.

More generally, a process of hierarchical merging, in which the present properties of any galaxy are determined by the necessarily haphazard details of its last major mergers, hardly seems consistent with the very high degree of organization revealed in this analysis. Hierarchical galaxy formation does not explain the commonplace gaseous galaxies we observe. So much organization, and a single controlling parameter—which cannot be identified for now—argue for some simpler model of formation. It would be illuminating to identify the single controlling galaxy parameter, but this cannot be attempted from the present data. To confuse correlation with cause is the critical mistake that can be made in this kind of analysis.

It is natural to ask why this fundamental line was not discovered before. To some extent it was because even the pioneers[24,25] and others[26–28] working with small numbers of optically selected spirals could reduce six observables to two; one relating to size, one to morphology. The strong optical selection effects, which hamper optical astronomers in detecting and measuring objects whose surface brightnesses are barely brighter than the sky[5,29], probably disguised galaxies' simplicity.

Received 27 March; accepted 26 August 2008.

1. de Jong, R. S. (ed.) *Island Universes: Structure and Evolution of Disk Galaxies* (Astrophys. Space Sci. Proc., Springer, 2007).
2. Baugh, C. M. A primer on hierarchical galaxy formation: the semi-analytical approach. *Rep. Prog. Phys.* **69**, 3101–3156 (2006).
3. Dalcanton, J. J., Spergel, D. N. & Summers, I. J. The formation of disk galaxies. *Astrophys. J.* **482**, 659–676 (1997).
4. Mo, H. J., Mao, S. & White, S. D. M. The formation of galactic discs. *Mon. Not. R. Astron. Soc.* **295**, 319–336 (1998).
5. Disney, M. J. in *The Low Surface Brightness Universe* (eds Davies, J. I., Impey, C. & Phillipps, S.) 9–18 (IAU Symp. 171, Astron. Soc. Pacif. Conf. Ser. 170, Astronomical Society of the Pacific, 1999).
6. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311**, 517–525 (1984).
7. Staveley-Smith, L. *et al.* The Parkes 21-cm multibeam receiver. *Publ. Astron. Soc. Austral.* **13**, 243–248 (1996).
8. Zwaan, M. A. *et al.* The HIPASS catalogue – II. Completeness, reliability, and parameter accuracy. *Mon. Not. R. Astron. Soc.* **350**, 1210–1219 (2004).
9. Meyer, M. J. *et al.* The HIPASS catalogue – I. Data presentation. *Mon. Not. R. Astron. Soc.* **350**, 1195–1209 (2004).
10. Adelman-McCarthy, J. K. *et al.* The fifth data release of the Sloan Digital Sky Survey. *Astrophys. J. Suppl. Ser.* **172**, 634–644 (2007).
11. Garcia-Appadoo, D. A., West, A. A., Dalcanton, J. J., Cortese, L. & Disney, M. J. Correlations among the properties of galaxies found in a blind HI survey. *Mon. Not. R. Astron. Soc*. (submitted); preprint at ⟨http://arxiv.org/abs/0809.1434⟩ (2008).
12. Strateva, I. *et al.* Color separation of galaxy types in the Sloan Digital Sky Survey imaging data. *Astron. J.* **122**, 1861–1874 (2001).
13. Kulkarni, S. R. & Heiles, C. in *Galactic and Extra-Galactic Radio Astronomy* 2nd edn (eds Kellerman, K. I. & Verschuur, G. L.) 95–153 (Springer, 1988).
14. Gavazzi, G., Pierini, D. & Boselli, A. The phenomenology of disk galaxies. *Astron. Astrophys.* **312**, 397–408 (1996).
15. Blanton, M. R. *et al.* The galaxy luminosity function and luminosity density at redshift $z = 0.1$. *Astrophys. J.* **592**, 819–838 (2003).
16. van der Kruit, P. The three-dimensional distribution of light and mass in disks of spiral galaxies. *Astron. Astrophys.* **192**, 117–127 (1988).
17. Haynes, M. P. & Giovanelli, R. Neutral hydrogen in isolated galaxies. IV – Results for the Arecibo sample. *Astron. J.* **89**, 758–800 (1984).
18. Rosenberg, J. L., Schneider, S. E. & Posson-Brown, J. Gas and stars in an HI-selected galaxy sample. *Astrophys. J.* **129**, 1311–1330 (2005).
19. Jolliffe, I. T. *Principal Component Analysis* (Springer, 1986).
20. Chatfield, C. & Collins, A. J. *Introduction to Multivariate Analysis* (Chapman & Hall, 1980).
21. Kormendy, J. & Kennicutt, R. C. Jr. Secular evolution and the formation of pseudobulges in disk galaxies. *Annu. Rev. Astron. Astrophys.* **42**, 603–683 (2004).
22. Zibetti, S. *et al.* 1.65 micron (H band) surface photometry of galaxies. I. Structural and dynamical properties of elliptical galaxies. *Astrophys. J.* **579**, 261–269 (2002).
23. Conselice, C. J. The fundamental properties of galaxies and a new galaxy classification system. *Mon. Not. R. Astron. Soc.* **373**, 1389–1408 (2006).
24. Brosche, P. The manifold of galaxies. Galaxies with known dynamical parameters. *Astron. Astrophys.* **23**, 259–268 (1973).
25. Balkowski, C. Statistical study of integral properties of galaxies measured in the 21-cm line. *Astron. Astrophys.* **29**, 43–55 (1973).
26. Bujarrabal, V., Guibert, J. & Balkowski, C. Multidimensional statistical analysis of normal galaxies. *Astron. Astrophys.* **104**, 1–9 (1981).
27. Shostak, G. S. Integral properties of late-type galaxies derived from HI observations. *Astron. Astrophys.* **68**, 321–341 (1978).
28. Whitmore, B. C. An objective classification system for spiral galaxies. I. The two dominant dimensions. *Astrophys. J.* **278**, 61–80 (1984).
29. Dalcanton, J. J., Spergel, D. N., Gunn, J. E., Schmidt, M. & Schneider, D. P. The number density of low-surface brightness galaxies with $23 < \mu_0$ 25 V Mag/arcsec^2. *Astron. J.* **114**, 635–654 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank the HIPASS team, and especially R. Ekers, A. Wright and L. Staveley-Smith of the Australian National Telescope at CSIRO Radiophysics in Sydney for their foresight and enterprise in getting the Multibeam project started. M.J.D. would like to thank M. Disney of the Geography Department at University College London for first pointing out the one-dimensional nature of this data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.J.D. (mjd@astro.cf.ac.uk).