nature astronomy

ARTICLES

https://doi.org/10.1038/s41550-021-01404-1

Check for updates

The origin of galaxy colour bimodality in the scatter of the stellar-to-halo mass relation

Weiguang Cui[1,2] ✉, Romeel Davé[1], John A. Peacock[1], Daniel Anglés-Alcázar[3,4] and Xiaohu Yang[2,5]

Recent observations reveal that, at a given stellar mass, blue galaxies tend to live in haloes with lower mass, while red galaxies live in more massive host haloes. The physical driver behind this is still unclear because theoretical models predict that, at the same halo mass, galaxies with high stellar masses tend to live in early-formed haloes, which naively leads to the opposite trend. Here, we show that the Simba simulation quantitatively reproduces the colour bimodality in the stellar-to-halo mass relation and reveals an inverse relationship between halo formation time and galaxy transition time. It suggests that the origin of this bimodality is rooted in the intrinsic variations of the cold gas content due to halo assembly bias. Simba's stellar-to-halo mass bimodality quantitatively relies on two aspects of its input physics: (1) jet-mode feedback from active galactic nuclei, which quenches galaxies and sets the qualitative trend, and (2) X-ray feedback from active galactic nuclei, which fully quenches galaxies and yields better agreement with observations. The interplay between the growth of cold gas and the quenching from active galactic nuclei in Simba results in the observed stellar-to-halo mass bimodality.

The galaxy stellar-to-halo mass relation (SHMR) represents a fundamental barometer for accretion and feedback processes in galaxy formation[1]. At low masses, the efficiency of star formation is low, increasing towards a peak of ~25% (the ratio between stellar mass and expected baryon mass) in $M_{\rm halo} \approx 10^{12} M_\odot$ haloes, and then dropping again at higher masses[2,3]. The origin of the SHMR has been explored extensively in physical models using semi-analytic[4–8] and hydrodynamic[9–11] techniques, which generally invoke star formation feedback to explain the SHMR below $M^\star$, and active galactic nucleus (AGN) feedback to explain the star formation inefficiency at higher halo masses[12].

Although a bimodality in galaxy colour has long been known[13–15], recently it has yielded new clues into the origin of the SHMR based on observations of a bimodality in the scatter around the mean SHMR at $z \approx 0$ (refs. [16–20]). At a given stellar mass, for reasonably massive systems (stellar mass $M_* \gtrsim 10^{10.2} M_\odot$, or halo mass $M_{\rm halo} \gtrsim 10^{11.5} M_\odot$), red (quenched) galaxies tend to live in more massive haloes, while blue (star-forming) galaxies occupy less massive haloes. A reinterpretation of this is that star-forming galaxies have a higher stellar mass than do quenched galaxies at a given halo mass. Note that binning in halo mass versus stellar mass may result in the so-called 'inversion problem'[21,22]; see Supplementary Information section 2 for further discussion.

Theoretical studies[23,24] broadly suggest that the scatter in the SHMR also correlates with halo formation time, which indicates the existence of assembly bias; that is, that galaxy properties at a given halo mass depend on halo formation time. However, linking galaxy colour to its host halo formation time is not straightforward, and contradictory results have been reported. A naive expectation is that early-formed haloes[25] would host early-formed galaxies, which would quench earlier and therefore be redder and less massive today[26]. However, observations suggest an opposite trend, which is not reproduced by theoretical models[22,23,27]. Therefore, the connection of the scatter in the SHMR to halo assembly time and central galaxy mass and colour is more complex than naively expected[17,28] (see Supplementary Information section 4 for further discussion), and galaxy evolution processes may play a critical role. For instance, processes that delay star formation without invoking overly strong supernova-driven outflows could explain the observed high M_*–$M_{\rm halo}$ ratios of blue central galaxies[29]. But which physical process(es) would enable this is not obvious.

Here, we seek to understand the cause of the SHMR bimodality using the state-of-the-art Simba simulation[30], which described in the Methods. We show that Simba quantitatively reproduces the observed SHMR bimodality for central galaxies, which is an important success. We examine its physical origin in terms of the transition time at which a galaxy goes from fast growing to slow growing. We demonstrate that the halo formation time drives a difference in the amount of cold gas within haloes, leading to halo assembly bias, while the transition time is driven by a declining cold gas content owing primarily to the onset of AGN feedback in Simba. This connects the SHMR colour bimodality back to its intrinsic physical origin in the physics of AGN feedback in Simba, and gives clues as to how such feedback must be triggered in order to quench galaxies as observed.

The colour bimodality in the SHMR

Simba well reproduces the observed bimodality in the SHMR relation, with values that are consistent with observations across the entire mass range probed, particularly when compared to the kinematic data (Fig. 1). The galaxy colour is separated according to ref. [16] (the Methods). We further refer to Supplementary Fig. 1 for the galaxy colour evolution timescale in Simba. At a given stellar mass, Simba shows that red galaxies live in larger haloes. Furthermore, Simba predicts that, at a given halo mass, blue galaxies have larger stellar masses, at least over the halo mass range of $10^{11.5} M_\odot \lesssim M_{\rm halo} \lesssim 10^{12.8} M_\odot$. This region is also free of the inversion problem[21,22] (Supplementary Fig. 2), as this is the interesting regime where quenching occurs. Simba's success in reproducing the observed SHMR colour bimodality motivates the investigation of the input physics that drives this agreement.

[1]Institute for Astronomy, University of Edinburgh, Edinburgh, UK. [2]Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China. [3]Department of Physics, University of Connecticut, Storrs, CT, USA. [4]Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA. [5]Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai, China. ✉e-mail: weiguang.cui@ed.ac.uk

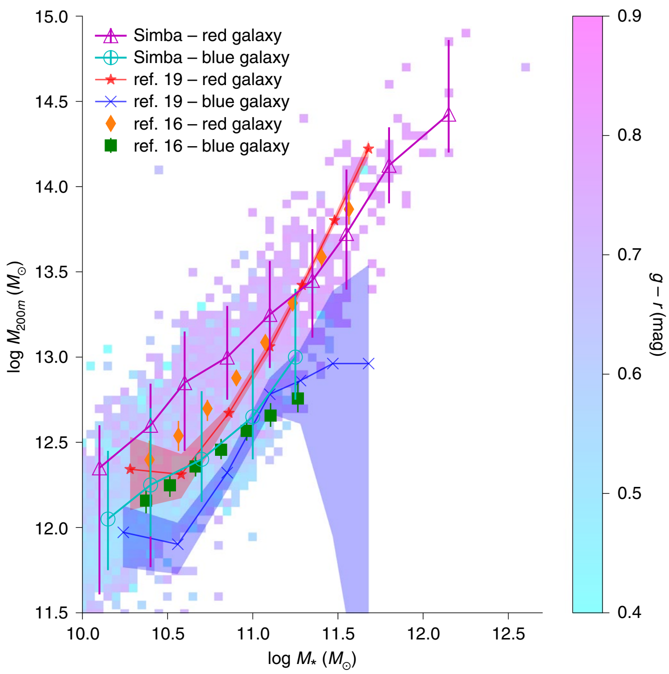

Fig. 1 | The SHMR at $z = 0$ from the Simba simulation, separated into red and blue galaxies, compared to observational results. The simulated galaxies are binned by galaxy and halo mass, with the median $g-r$ colour in each bin shown by the colour bar. The open cyan circles and magenta triangles show median values for blue and red galaxies, respectively, with error bars spanning the 16th to 84th percentiles. Using a separation based on specific star formation rate (sSFR, $\log[\mathrm{sSFR}(\mathrm{Gyr}^{-1})] = -2$) yields a similar result. Observational results with halo masses based on weak lensing (ref. [19]) and satellite kinematics (ref. [16]) are shown with symbols and colours indicated in the legend. For these, both the shaded regions and error bars indicate 1σ errors. Note that the same halo mass definition is adopted as in observations (M_{200m}, the mass enclosing 200 times the mean density), and we correct for minor differences in cosmological parameters.

We show in Fig. 2 the same SHMR as in Fig. 1, but now colour coded by halo formation time (z_F) in the left panel, and galaxy transition time (z_T) in the right panel. Halo formation time z_F is defined as the redshift at which the halo's most massive progenitor was half the present-day halo mass. Galaxy transition time z_T is defined as when the central galaxy's specific stellar growth rate exceeds 10 Gyr^{-1}, and characterizes the time at which the galaxy transitions from a stellar growth mode to a quiescent mode (the Methods and Supplementary Fig. 3). Note that here and in all subsequent figures except the last, a friends-of-friends (FoF) halo mass M_{halo} instead of M_{200m} is used (the Methods). Finally, Supplementary Information section 2 has a more detailed discussion of the inversion problem on z_F and z_T, but this does not influence our results in the key mass range.

Late-formed haloes tend to have higher halo mass at a given M_*, or conversely, at a given M_{halo}, early-formed haloes host larger galaxies (Fig. 2a). Therefore, Simba yields the same halo assembly bias trend as seen in previous results[23,24,27,31]. Although it is reassuring that models broadly agree on the correlation between the scatter in galaxy stellar mass and halo formation time, this does not identify the underlying physical driver(s) of the colour bimodality. Indeed, theoretical models tend to predict galaxies with high stellar mass at a given halo mass to have a red colour[22,27] (Supplementary Information section 4). To understand the colour bimodality in the SHMR, we must connect the halo formation to the galaxy's mass accretion history, which sets its colour.

We parameterize the galaxy mass accretion history by the galaxy transition time z_T. In the right panel of Fig. 2, we re-display the SHMR from Simba, now with the galaxies colour coded by z_T. Clearly, z_T is also strongly correlated with the scatter in the SHMR. Linking to galaxy colour bimodality, we find that blue galaxies lie in the region of the SHMR parameter space that is occupied by late-transition galaxies ($z_T < 1$), while red galaxies tend to occupy the region where early-transition galaxies ($z_T \geq 1$) are found. For more details on the connection between galaxy colour and its transition time, see Supplementary Fig. 1.

A key point is that the galaxy transition time is anti-correlated with the halo formation time: galaxies that transition later have a lower halo mass at a given M_*, which implies that early-transition galaxies tend to live in late-formed haloes, while late-transition galaxies are in early-formed haloes at a given halo mass. This shows that haloes that assemble quickly tend to host bluer galaxies than haloes that assemble more slowly. This inverse relationship between z_F and z_T gives rise to stellar population downsizing, in which massive haloes that assembled most recently contain the oldest stellar populations (early transition). Connecting with the observed colour bimodality, blue (late-transition) galaxies tend to live in haloes with a lower mass at a given stellar mass, thus in haloes with earlier formation times. Note that, at $M_{halo} \gtrsim 10^{13} M_\odot$, the early-transition galaxies tend to have a higher stellar mass compared to the late-transition ones (Supplementary Fig. 2). At lower halo mass that is beyond the scope of this Article, this relation may be inversed as suggested by ref. [32]. The inverse relation between z_T and z_F reveals how galaxy colour is connected to halo assembly bias and thus to SHMR bimodality, but it does not pinpoint the physical driver(s) in Simba. Towards this, next we investigate the evolution of the SHMR.

The evolution of the SHMR with mass growth histories

To see how the halo and galaxy growth (see Supplementary Fig. 3 for an example) manifest themselves on the SHMR plane, Fig. 3 shows the median evolutionary tracks of the SHMR separated into different halo mass bins and three colour regions (defined in Fig. 2). Here, we invert the axes to illustrate our points more clearly. We note that, even in the quiescent state well after transition, galaxies can still grow by mergers and residual star formation. Although we do not explicitly separate the in-situ and ex-situ growth when computing the transition time, previous work has shown that the fast stellar growth phase is dominated by in-situ growth, while ex-situ growth becomes important only for massive quiescent galaxies[33,34].

Prior to the transition time, the slopes of the tracks in the same halo mass bin appear anti-correlated with the final stellar masses—the lower the final M_*, the higher the slope. The low slope indicates fast halo growth relative to stellar growth. If we assume that halo mass growth concurrently brings in gas, fast halo growth will bring more gas at high redshift to fuel star formation. We speculate that, owing to high early cosmic densities, this gas is expected to be cool[35,36]. We will show that this enables the central galaxy to sustain its star formation for longer, which results in a higher stellar mass compared to that of galaxies at the same final halo mass that have a steeper early slope.

If we compare the curves that result in a similar final stellar mass but are in different halo mass bins, it seems that they share a similar growth trend at high redshift and reach the galaxy transition time at a similar halo mass. However, the galaxies that end up in red regions have a much earlier transition time than those in blue or green regions. Therefore, after the transition time, they have a longer time to grow their halo mass, while their stellar mass grows more slowly, depositing them in the lower-right portion of this SHMR plane. Conversely, galaxies that end up in the blue or green regions have a later transition time. This leaves less time for them to grow their halo, and they finish in the upper right (shown as the blue region) or middle (green region) of the SHMR plane.

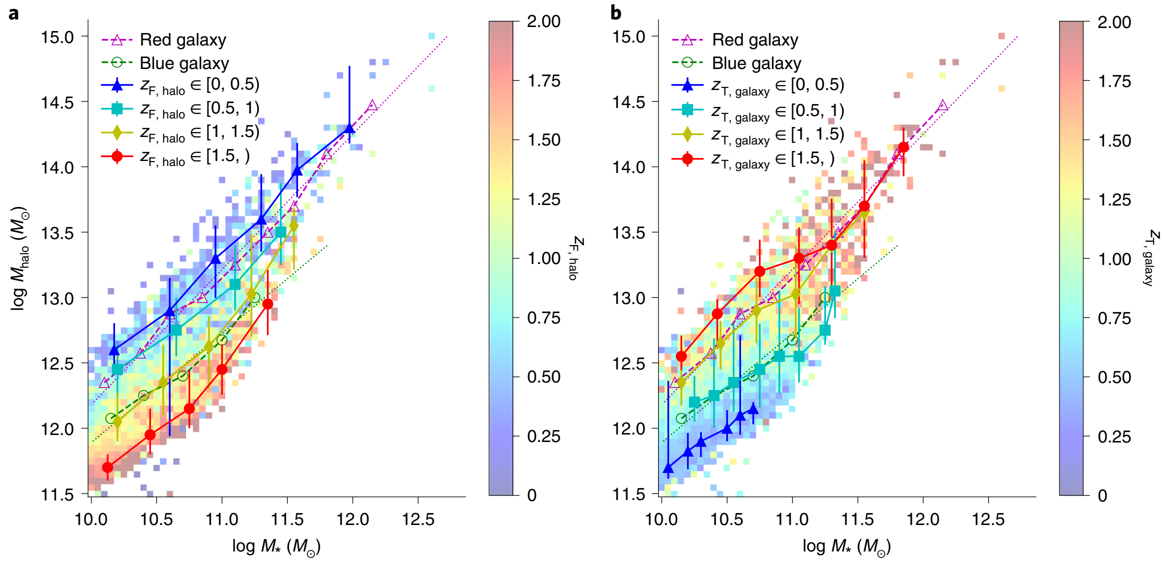

Fig. 2 | The SHMR from the Simba simulation. a,b, The SHMR colour coded by the median halo formation redshift (**a**) and by galaxy transition redshift (**b**). Lines of different colours show median values for the transition and formation redshift ranges indicated in the legends, with error bars showing 16th to 84th percentiles. Median lines for red and blue galaxies shown in the same symbols and colours as in Fig. 1 are also included for comparison, with dotted magenta and cyan lines showing the best linear fits to those relations: $\log M_{halo} = 0.97\log M_* - 1.77$ for red galaxies, and $\log M_{halo} = 1.2\log M_* - 4.23$ for blue.

As a side note, in Fig. 3 the medians in the blue region still show some post-transition growth in stellar mass, which seemingly contradicts our definition of a transition going into slow growth. This growth is driven by a small number of galaxies that happen to have fairly high specific star formation rates, such that they are close to our transition limit of 0.1. Using a lower threshold would bring the transition time closer to $z = 0$, but it would not qualitatively change our conclusions; for simplicity, we keep the existing threshold for this work.

Figure 3 has interesting implications for the driver of galaxy quenching. If we assume the transition time is correlated with the quenching time (albeit later on, because galaxies are still fairly gas rich at transition), then our results suggest that a fixed halo mass cannot be the sole driver of quenching, as has been claimed[37,38], at least for halo masses of $11.5\,M_\odot \lesssim \log M_{halo} \lesssim 12.8\,M_\odot$, where Simba's SHMR does not suffer from inversion issues. This is evident because the halo masses at a given transition time for central galaxies are spread out over an order of magnitude and do not show a single characteristic halo mass at which a galaxy enters its quiescent state. A similar conclusion that the halo mass is not a direct cause of galaxy quenching is reached by investigating the M_{HI}–M_{halo}–M_* relations[39]. However, halo mass must still play a key role, given that central galaxies in higher mass haloes have an earlier transition time and that the fraction of quenched galaxies in high-mass haloes is higher. If the characteristic halo mass is still connected with galaxy quenching, one possible solution is that this halo quenching mass is redshift dependent; that is, there is a higher characteristic halo mass for quenching at higher redshifts[40,41]. The anti-correlation between z_T and z_F and its role in setting the scatter in the SHMR indicates that the halo mass growth history plays a key role in shaping the SHMR. In the next section, we bring in the next piece of this puzzle, the cold gas evolution, as a driver of this anti-correlation.

The origin of the colour bimodality

As shown in the previous section, at early epochs all galaxies are fast growing and lie along a similar SHMR track. We will demonstrate that, in a given (final) halo mass bin, early-forming haloes collect copious amounts of cool gas, which then begins baryon cycling via

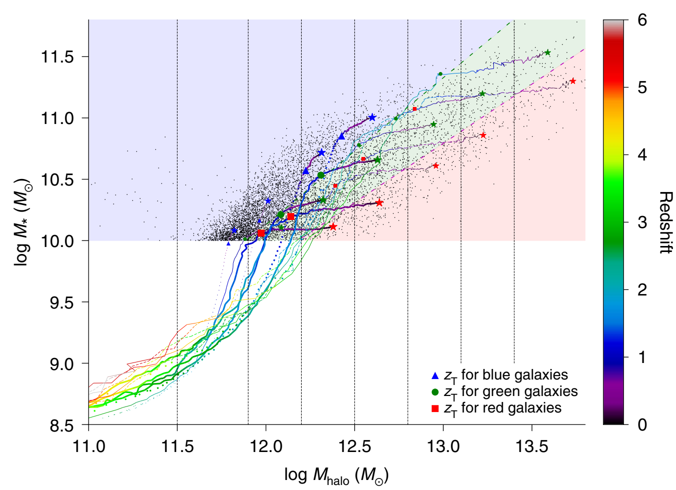

Fig. 3 | The evolution track of the SHMR. The SHMR, shown with M_{halo} on the x axis and M_* on the y axis. Galaxies at $z = 0$ (black dots) are separated into three different regions (blue, green and red) by the median red and blue galaxy best-fit lines from Fig. 2, and into various halo mass bins marked by the vertical dotted lines. The median stellar mass and halo mass in each zone is marked with a star symbol: blue for those lying above the fitting line for blue galaxies (early-formed halo with late-transition galaxies), red for those below the fitting line for red galaxies (late-formed halo with early-transition galaxies) and green for those in between. The median SHMR evolution track of each zone is shown by the coloured curves (dotted for the blue region, solid for the green region and dashed for the red region), with the redshift shown in the colour bar. The median galaxy transition times are marked with the symbols shown in the legend. We consider galaxies in the 15 zones with $M_{halo} > 10^{11.5}\,M_\odot$. Note that here we use the inverted SHMR plane, putting M_{halo} on the x axis, so that we can avoid conflating with the halo formation time dependence by keeping halo mass as the independent variable. We further use thick lines to highlight two interesting halo mass regions: $12.2 < \log M_{halo} \leq 12.5$ and $12.5 < \log M_{halo} \leq 12.8$, which each include all three colour regions.

outflows to continue to provide fuel for galaxy stellar growth into later epochs—this yields a late transition time. Conversely, galaxies in late-forming haloes tend to have an earlier transition time, which is driven by a lower cold gas supply and exacerbated by AGN jet feedback that first appears at low black-hole accretion rates (the Methods). Furthermore, once the transition time is reached, galaxies grow primarily in halo mass with only meagre stellar mass growth, separating these red galaxies from the SHMR relation of the star-forming systems (Fig. 3). To support this scenario, first we need to examine the early cold gas content difference that results from halo assembly bias. It turns out that this will implicate Simba's AGN feedback as a key actor in SHMR bimodality.

The intrinsic origin: halo assembly bias and cold gas content. Figure 4 shows the evolution of the cold gas fraction in the six regions in the SHMR plane, as depicted in Fig. 3; here, we only consider the middle two $z=0$ halo mass bins, where galaxies exist in all three coloured stellar mass regions and are also free from the inversion problem. The cold gas fraction is defined as the H I + H_2 gas mass in the central galaxy divided by the host halo mass (Supplementary Information section 3).

The galaxies in the blue region are hosted by early-formed haloes; that is, they are more massive in halo mass at high redshifts than are their present-day counterparts. We speculate that, owing to a higher early accretion rate, this higher halo mass results in a higher gas density that enhances cooling, thereby plateauing at a higher gas fraction at intermediate redshifts[42] when the galaxy settles into an equilibrium between inflows, outflows and star formation[43]. This higher gas fractions persists through the transition time (marked as a large square), as shown in Fig. 4. We suggest that this sustains the central galaxy's star formation, but also yields a high Eddington ratio for the central black hole, which in Simba keeps the AGN feedback in a milder radiative mode that does not quench the galaxy. The enhanced gas content at a given halo mass, driven by the early halo assembly, persists all the way to $z=0$, growing galaxies larger and keeping them bluer.

By contrast, galaxies in the red region with lower cold gas fractions are more easily able to reach the Eddington ratio limit of a few percent, triggering the jet-mode AGN feedback in Simba. Jet-mode feedback has an order-of-magnitude higher wind speed than the radiative-mode AGN feedback, and it is commensurately more energetic. We suggest that this will result in notably lower halo gas contents overall, and that the gas that remains is mostly hot[30,44,45]. The depleted hot halo gas cannot fuel star formation, leaving such late-formed haloes with redder, less massive galaxies. Furthermore, the cold gas fractions are higher in the lower halo mass bin for all three regions, as higher halo masses host more massive galaxies with weaker winds that result in more effective conversion of gas into stars[46].

The physical driver of galaxy transition: jet-mode AGN feedback. It is clear that the cold gas supply in the halo is an important driver of the SHMR bimodality. To quantify these effects and connect them back to the input physics in Simba, we introduce a new quantity, the galaxy crossing time, defined as the time when $M_* = M_{gas}$. We correlate this with the time of AGN jet turn-on, which in Simba is when the black hole's Eddington ratio drops to below 2% (the Methods and Supplementary Information section 3), as well as with the transition time t_T (which corresponds to z_T).

We show the correlations between these various times in Fig. 5. In the left panel, we show the crossing time $t_{M_*=M_{gas}}$ as a function of the galaxy transition time (t_T), while in the right panel we show $t_{jet\ on}$ versus t_T. Remarkably, both $t_{M_*=M_{gas}} - t_T$ and $t_{jet\ on} - t_T$ follow a nearly one-to-one relation, which confirms the driving role of these aspects in setting the galaxy transition time. As an aside, note the ~0.5 Gyr delay between $t_{M_*=M_{gas}}$ and the transition time

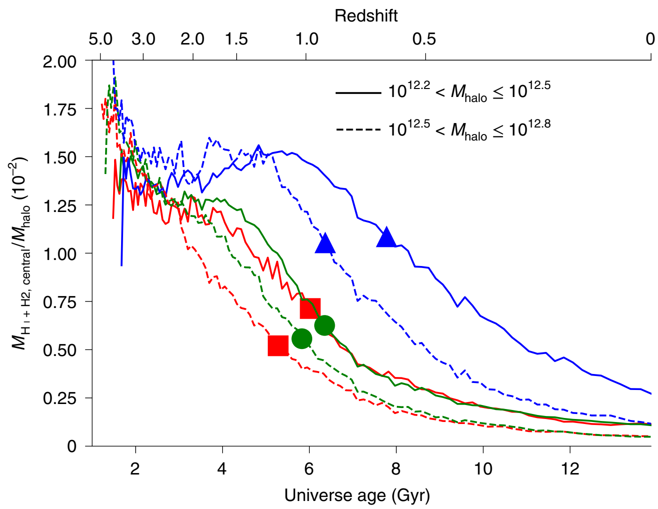

Fig. 4 | The evolution of the gas fraction. The median cold gas fraction, defined as the central galaxy neutral gas mass divided by the halo mass, versus cosmic time. Two selected halo mass bins are shown as indicated in the legend: solid lines are for the lower halo mass bin, and dashed lines are for the higher halo mass bin shown in Fig. 3. Different coloured lines depict results in the corresponding coloured regions in Fig. 3. The median galaxy transition times are marked as blue triangles for blue galaxies, green circles for green galaxies and red squares for red galaxies along the curves.

for massive haloes (red and yellow dashed lines), in agreement with Supplementary Fig. 3, which indicates that galaxy quenching requires a lower gas fraction than does the galaxy transition. This correlation is in agreement with the expulsion of efficiently cooling gas from the circumgalactic medium as a crucial step in quenching a galaxy[47]. Although the galaxy star formation rate correlates with the cold gas mass[48], the fact that t_T is proportional to $t_{M_*=M_{gas}}$ is a non-trivial consequence of the upward trend of increasing cold gas mass at early times, which is tightly connected with halo formation time, and the downward trend of decreasing cold gas mass at late times, which can be driven by depletion by star formation, heating or outflows by feedback, and/or virial shock heating[35]. Crucially, the lower gas fraction connects to a low Eddington ratio that triggers the jet-mode AGN feedback, as shown in the right panel of Fig. 5, which quenches the galaxy[30].

We have demonstrated that Simba's success in reproducing the SHMR scatter is owing to its quenching feedback that coincides with the halo cold gas content dropping to low values, therefore triggering the jet-mode AGN feedback that transitions the galaxy. However, this does not fully explain the SHMR bimodality, as we show next; one additional ingredient is needed in Simba.

Enhancing the galaxy colour separation: X-ray feedback. We can quantitatively assess how Simba arrives at its SHMR bimodality by examining Simba's AGN feedback variant runs. The left panel of Fig. 6 shows the SHMR as in Fig. 1 from three 50 h^{-1} Mpc runs with identical initial conditions: a full Simba physics run (fiducial), a run turning off X-ray AGN feedback ('No-X') and a run turning off both X-ray and jet-mode feedback ('No-jet'). Note that M_{200m} instead of M_{FOF} halo masses are used here for a consistent comparison with the observed results. The fiducial run shows similar results to the main 100 h^{-1} Mpc Simba run in Fig. 1, which confirms that our results are not sensitive to simulation volume.

Starting with the No-jet case, we see that the median SHMR for blue galaxies is similar to the fiducial full-physics run, but the SHMR for red galaxies is drastically shifted towards higher M_*, so that the

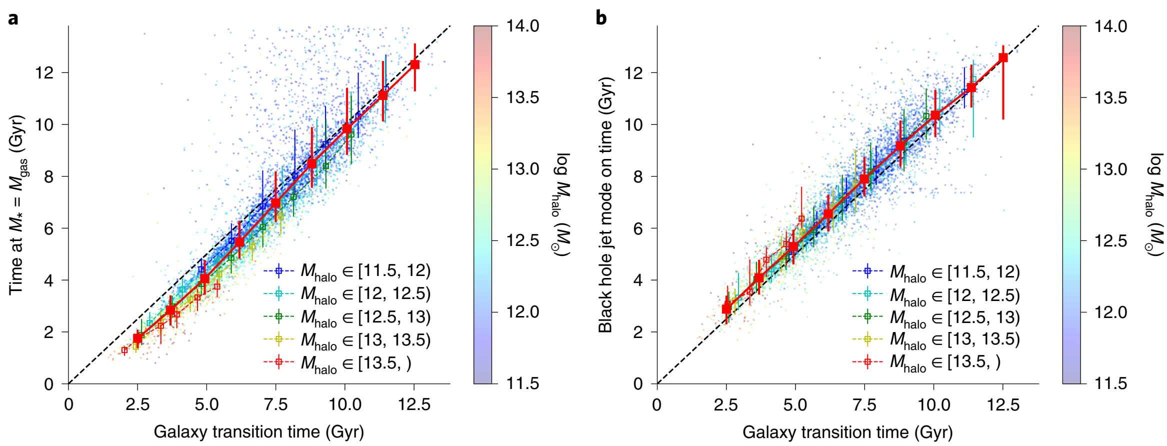

Fig. 5 | The correlation between galaxy transition time with crossing time and AGN jet-on time. a, Galaxy crossing time, when the galaxy stellar mass is equal to its cold gas mass, versus galaxy transition time when it enters a slowly growing phase. **b**, Black-hole jet-on time, when jet-mode AGN feedback turns on, versus galaxy transition time. Galaxies are colour coded by their halo mass at $z = 0$ (in units of log[M_{halo} ($M_\odot$)]), with the medians of different halo mass bins in coloured dashed lines. The black dashed line indicates the 1-to-1 relation. The red line with error bars shows the median values of all haloes and 16th to 84th percentiles after binning in galaxy transition time.

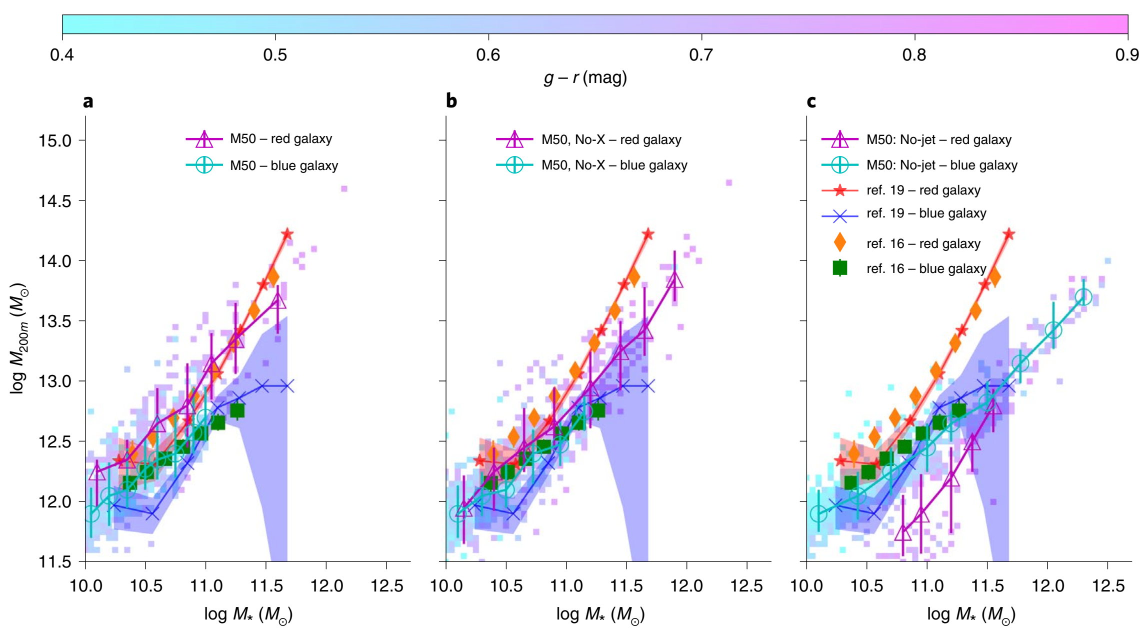

Fig. 6 | The SHMR of three simulations with different baryon models. a–c, The SHMR subdivided into red and blue galaxies by the same colour cut as in Fig. 1, but for a smaller 50 h^{-1} Mpc simulation box (M50) with the full Simba feedback (the same as the default run) in an identical run (**a**), a 'No-X' run turning off only X-ray AGN feedback (**b**) and a 'No-jet' run additionally turning off the jet-mode AGN feedback (**c**). Note that the notations and error bars in Fig. 1 are adopted here.

colour bimodality in the SHMR is now reversed! In this way, without quenching, the SHMR follows the naive expectation that galaxy growth simply follows halo growth, so late-formed haloes continue to harbour star formation and grow their stellar mass until today. This is in clear disagreement with observations.

Turning on only jet feedback (No-X), again the blue SHMR remains mostly unchanged, but the red galaxies' SHMR now has moved slightly above the blue. Therefore, jet feedback goes most of the way towards establishing the SHMR colour bimodality—but it still does not fully match observations. This requires turning on X-ray AGN feedback, which fully quenches galaxies[30], and moves the quenched galaxy SHMR to somewhat lower M_* and thus quantitatively reproduces the observations.

The basic qualitative trend of the SHMR colour bimodality in Simba is owing to jet-mode AGN feedback, while quantitative agreement further requires the inclusion of X-ray AGN feedback. This completes the connection between the SHMR scatter bimodality and the driving input physics in Simba.

Conclusion

We recap our scenario for the origin of the colour bimodality in the SHMR in Simba, which successfully reproduces observations without any specific tuning. The correlation between the scatter in the SHMR and halo formation time reveals that the colour bimodality is rooted in the intrinsic variation of the early cold gas content owing to halo assembly bias: at a given final halo mass, early-formed haloes have higher cold gas fractions, and vice versa. In late-formed haloes with low cold gas fractions, the galaxy transitions earlier, setting up an anti-correlation of the halo assembly time and galaxy transition time. The physics driving this in Simba is primarily jet-mode AGN feedback, which appears at low Eddington ratios that occur when cold gas fractions are low, and quenches the galaxy. Furthermore, the X-ray AGN feedback model in Simba, which begins for jet-mode galaxy with a low gas fraction, removes any residual star formation and therefore keeps the galaxy in quiescence, leading to a red colour. Once the galaxy transitions to a slow-growing phase, the halo continues to grow, placing these red galaxies in more massive haloes by $z=0$. Conversely, early-formed haloes accumulate plentiful cold gas early on, which delays the onset of quenching and sustains star formation for longer to yield bluer galaxies.

We have used the Simba simulation to illustrate how its input physics drives the SHMR bimodality. Even if the details of Simba's feedback models are not correct, our results are likely to be generalizable to other galaxy formation models. What Simba suggests is that, to obtain the SHMR colour bimodality, it is necessary that early-forming haloes obtain more cold gas (which is expected from structure formation), and that they must also retain this advantage to $z=0$. To do this, we suggest that the feedback that quenches galaxies, whatever that is (for example, jet feedback in Simba), must operate preferentially in haloes with lower cold gas fractions. This is critical for producing the observed trend of the SHMR bimodality. Obtaining quantitative agreement with data requires keeping galaxies fully quenched and red; in Simba, this occurs by X-ray feedback that pushes residual cold gas out. Although other models may implement quenching feedback in different ways, our results suggest that if quenching is connected to haloes that have preferentially lower amounts of cold gas to fuel star formation, the net effect is likely to be similar to that of Simba's AGN feedback modules, and will qualitatively reproduce the observed SHMR colour bimodality. That said, our results suggest that obtaining quantitative agreement with the SHMR colour bimodality provides a stringent test for models of galaxy quenching, particularly as constraints improve with upcoming facilities such as the Vera Rubin Telescope.

Methods

The Simba simulation. The 100 h^{-1} Mpc main Simba simulation[30] is used for this study. Simba is based on the MUFASA simulation[49] with its sub-resolution star formation and stellar feedback prescriptions, with galactic wind scaling taken from the FIRE simulations[50,51]. Simba further includes two models of black-hole growth prescriptions: the torque-limited accretion model from cold gas[52,53], and Bondi accretion from hot gas. AGN feedback is modelled by kinetic bipolar outflows, the strength of which depends on the black-hole accretion rate, separated into three modes: a 'radiative mode' at high Eddington ratio to drive multi-phase winds at velocities of ~10^3 km s^{-1}, a 'jet mode' at low Eddington ratios ($f_{\rm Edd}<0.2$), at which AGNs drive hot gas in collimated jets at velocities of ~10^4 km s^{-1}, and X-ray heating from black holes, which aims to represent the momentum input from hard photons radiated off from the accretion disk[54], and which is only initiated after jet-mode feedback is turned on. This simulation is tuned to reproduce the stellar mass function up to redshift $z\approx6$. Furthermore, it reproduces the quenched fractions and the main sequence of star-forming galaxies at various M_* (refs. [30,55]), the observed black-hole mass – galaxy stellar mass and velocity dispersion relations[56], the observed dust mass function at $z=0$ (ref. [57]), and many other galaxy properties[58].

In addition, a series of smaller box (50 h^{-1} Mpc) simulations with the same resolution as the 100 h^{-1} Mpc main Simba simulation, but with different feedback models, are also used to investigate the effects of different AGN feedback processes in this study: a fiducial run that shares everything with the 100 h^{-1} Mpc run, a 'No-X' run that turns off only X-ray feedback and a 'No-jet' run that turns off both X-ray and jet-mode feedback. Note that these 50 h^{-1} Mpc simulations also share the same initial condition. By comparing these models, we can isolate the impact of individual AGN feedback modes on the SHMR.

Haloes are identified on the fly during the simulation run using a three-dimensional FoF algorithm with a linking length of 0.2 times the mean inter-particle spacing. The halo masses in this paper are the FoF halo mass, except for those in Figs. 1 and 6, for which we compute M_{200m} to be consistent with observational comparisons; this makes little difference to our results. The adoption of the FoF halo mass is to avoid pseudo halo growth introduced by the overdensity method[59], which is important for the study of evolutionary trends in the section 'The evolution of the SHMR with mass growth histories'. However, there is little difference between the two mass definitions at $z=0$, as shown by the negligible differences between the red and blue galaxy median lines in Figs. 1 and 2.

Galaxies are identified using a six-dimensional phase-space galaxy finder within each dark matter halo in the yt-based package Caesar. Many galaxy properties are computed after the identification. The galaxy magnitudes in the g and r bands of the Sloan Digital Sky Survey are computed using the PyLoser package within Caesar, which employs the Flexible Stellar Population Synthesis code[60,61] to compute spectral energy distributions and includes line-of-sight extinction based on the self-consistently evolved dust content in Simba. We apply a stellar mass cut of $M_*\geq10^{10}M_\odot$ to ensure that we include only well-resolved central galaxies ($\gtrsim$500 star particles), which in any case covers the same mass range as the observations shown in Fig. 1. The simulated galaxies are separated further into red and blue by their $g-r$ colour, as in observations[16]. We note that the separation line, $(g-r)_{\rm cut}=0.65+0.075\times(\log(M_*/[h^2_{0.71}M_\odot])-10)$, has a slightly lower slope compared to that in ref. [16], because massive galaxies in Simba include the intra-cluster light, which has a younger age than the brightest cluster galaxy[62]. Finally, galaxies and haloes are linked with their progenitors through matching their unique particle identification numbers. As we focus only on the central galaxy and its host halo in this study, the merger history is built on the basis of the main progenitors of $z=0$ haloes. If not specified, our results apply only within this resolved mass range ($M_*\gtrsim10^{10}M_\odot$, or $M_{\rm halo}\gtrsim10^{11.5}M_\odot$).

Halo formation, galaxy transition and jet-mode AGN feedback ignition times. As illustrated in Supplementary Fig. 3, the histories of galaxy and halo formation are rather different, and the black-hole growth history is more similar to the galaxy growth history. Individual haloes or galaxies have fairly similar growth histories as the medians. We use the commonly adopted half-mass redshift, when the halo's most massive progenitor obtained half of its $z=0$ mass, for the halo formation time. We have calculated the halo formation time both by interpolation of the data points and from the data smoothed with the Savitzky–Golay filter[63], which has been integrated into SciPy[64,65]. The two methods give a consistent result. Galaxy mass accretion histories generally can be characterized by two processes: a fast-growing period in its early phase and a constant or quiescent period after the transition. Therefore, it is natural to choose the connecting point between the two periods as their transition time. First, we fit the galaxy formation history with a step function by joining an error function term (for the fast-growing period) and a linear term (for the constant period). We note that we simply use the total stellar mass of galaxies as they evolve without considering in-situ or ex-situ growth modes or rejuvenation processes. The advantages of this simple treatment are that we do not need a separate explanation for the scatter in the two phases and that we avoid making the difficult separation of in-situ and ex-situ growth in observations. Then, we get the galaxy transition time z_T through the slope of the fitting curve, as the time when

$$\frac{{\rm d}\log M_*}{{\rm d}t({\rm Gyr})}<0.1 \quad (1)$$

We note here that this definition is similar to a threshold in the specific star formation rate, but it is based on a much longer time baseline, and it includes stellar mass brought in by mergers. Our conclusions are not affected by the choice of this threshold, which only produces a systematic shift in the transition time.

The black-hole mass growth history also can be roughly separated into a fast-growing period in its early phase and a constant period later on. This is consistent with two modes of black-hole accretion in Simba. Furthermore, this also correlates with the radiative and jet-mode AGN feedback, which is dominated in high Eddington ratios in the fast black-hole mass growing period and in low Eddington ratios ($f_{\rm Edd}<0.2$) in the constant period, respectively. To directly link the black-hole mass change with the Eddington ratio, we use equation 11 from ref. [56] to correlate the Eddington ratio with the black-hole specific growth rate, and define the jet-mode feedback ignition time as the time when

$$\frac{{\rm d}\log M_\bullet}{{\rm d}t({\rm Gyr})}<0.9 \quad (2)$$

Data availability

The Simba simulation snapshots with Caesar halo and galaxy catalogues are publicly available at http://simba.roe.ac.uk/. The processed source data for producing the figures in this paper are available at the author's repository: https://bitbucket.org/WeiguangCui/ms-mhalo-scatter/src/master/.

Code availability

The Simba simulation is run with a private version of Gizmo, which is available from the corresponding author upon reasonable request. The galaxy catalogue of the Simba simulation is produced by Caesar, which is publicly available at https://github.com/dnarayanan/caesar. The detailed analysis pipeline scripts are available in the author's repository: https://bitbucket.org/WeiguangCui/ms-mhalo-scatter/src/master/.

Received: 19 April 2020; Accepted: 21 May 2021;
Published online: 5 July 2021

References

1. Wechsler, R. H. & Tinker, J. L. The connection between galaxies and their dark matter halos. *Annu. Rev. Astron. Astrophys.* **56**, 435–487 (2018).
2. Moster, B. P., Naab, T. & White, S. D. M. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *Mon. Not. R. Astron. Soc.* **428**, 3121–3138 (2013).
3. Behroozi, P. S., Wechsler, R. H. & Conroy, C. The average star formation histories of galaxies in dark matter halos from $z=0$–8. *Astrophys. J.* **770**, 57 (2013).
4. Croton, D. J. et al. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365**, 11–28 (2006).
5. Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E. & Hernquist, L. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **391**, 481–506 (2008).
6. Guo, Q. et al. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *Mon. Not. R. Astron. Soc.* **413**, 101–131 (2011).
7. Henriques, B. M. B. et al. Galaxy formation in the Planck cosmology I. Matching the observed evolution of star formation rates, colours and stellar masses. *Mon. Not. R. Astron. Soc.* **451**, 2663–2680 (2015).
8. Croton, D. J. et al. Semi-analytic galaxy evolution (SAGE): model calibration and basic results. *Astrophys. J. Suppl. Ser.* **222**, 22 (2016).
9. Vogelsberger, M. et al. Introducing the Illustris project: simulating the coevolution of dark and visible matter in the Universe. *Mon. Not. R. Astron. Soc.* **444**, 1518–1547 (2014).
10. Schaye, J. et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. R. Astron. Soc.* **446**, 521–554 (2015).
11. Pillepich, A. et al. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Mon. Not. R. Astron. Soc.* **475**, 648–675 (2018).
12. Somerville, R. S. & Davé, R. Physical models of galaxy formation in a cosmological framework. *Annu. Rev. Astron. Astrophys.* **53**, 51–113 (2015).
13. Strateva, I. et al. Color separation of galaxy types in the Sloan Digital Sky Survey imaging data. *Astron. J.* **122**, 1861–1874 (2001).
14. Baldry, I. K. et al. Quantifying the bimodal color–magnitude distribution of galaxies. *Astrophys. J.* **600**, 681–694 (2004).
15. Balogh, M. L. et al. The bimodal galaxy color distribution: dependence on luminosity and environment. *Astrophys. J. Lett.* **615**, L101–L104 (2004).
16. More, S. et al. Satellite kinematics III. Halo masses of central galaxies in SDSS. *Mon. Not. R. Astron. Soc.* **410**, 210–226 (2011).
17. Wojtak, R. & Mamon, G. A. Physical properties underlying observed kinematics of satellite galaxies. *Mon. Not. R. Astron. Soc.* **428**, 2407–2417 (2013).
18. Velander, M. et al. CFHTLenS: the relation between galaxy dark matter haloes and baryons from weak gravitational lensing. *Mon. Not. R. Astron. Soc.* **437**, 2111–2136 (2014).
19. Mandelbaum, R. et al. Strong bimodality in the host halo mass of central galaxies from galaxy–galaxy lensing. *Mon. Not. R. Astron. Soc.* **457**, 3200–3218 (2016).
20. Posti, L. & Fall, S. M. Dynamical evidence for a morphology-dependent relation between the stellar and halo masses of galaxies. *A&A* **649**, 119–135 (2021).
21. Moster, B. P., Naab, T. & White, S. D. M. EMERGE – an empirical model for the formation of galaxies since $z \sim 10$. *Mon. Not. R. Astron. Soc.* **477**, 1822–1852 (2018).
22. Moster, B. P., Naab, T. & White, S. D. M. EMERGE – empirical constraints on the formation of passive galaxies. *Mon. Not. R. Astron. Soc.* **499**, 4748–4767 (2020).
23. Zehavi, I. et al. The impact of assembly bias on the galaxy content of dark matter halos. *Astrophys. J.* **853**, 84 (2018).
24. Matthee, J. et al. The origin of scatter in the stellar mass–halo mass relation of central galaxies in the EAGLE simulation. *Mon. Not. R. Astron. Soc.* **465**, 2381–2396 (2017).
25. Wang, H., Mo, H. J., Jing, Y. P., Yang, X. & Wang, Y. Internal properties and environments of dark matter haloes. *Mon. Not. R. Astron. Soc.* **413**, 1973–1990 (2011).
26. Lim, S. H., Mo, H. J., Wang, H. & Yang, X. An observational proxy of halo assembly time and its correlation with galaxy properties. *Mon. Not. R. Astron. Soc.* **455**, 499–510 (2016).
27. Matthee, J. & Schaye, J. The origin of scatter in the star formation rate–stellar mass relation. *Mon. Not. R. Astron. Soc.* **484**, 915–932 (2019).
28. Zu, Y. et al. Does concentration drive the scatter in the stellar-to-halo mass relation of galaxy clusters? Preprint at https://arxiv.org/abs/2012.08629 (2020).
29. Rodríguez-Puebla, A. et al. The stellar-to-halo mass relation of local galaxies segregates by color. *Astrophys. J.* **799**, 130 (2015).
30. Davé, R. et al. SIMBA: cosmological simulations with black hole growth and feedback. *Mon. Not. R. Astron. Soc.* **486**, 2827–2849 (2019).
31. Tojeiro, R. et al. Galaxy and mass assembly (GAMA): halo formation times and halo assembly bias on the cosmic web. *Mon. Not. R. Astron. Soc.* **470**, 3720–3741 (2017).
32. Feldmann, R., Faucher-Giguère, C.-A. & Kereš, D. The Galaxy-Halo connection in low-mass halos. *Astrophys. J.* **871**, L21 (2019).
33. Hirschmann, M. et al. The stellar accretion origin of stellar population gradients in massive galaxies at large radii. *Mon. Not. R. Astron. Soc.* **449**, 528–550 (2015).
34. Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S. & Behroozi, P. Physical correlations of the scatter between galaxy mass, stellar content and halo mass. *Mon. Not. R. Astron. Soc.* **493**, 337–350 (2020).
35. Kere, D., Katz, N., Weinberg, D. H. & Dave, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
36. Dekel, A. et al. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).
37. Zu, Y. & Mandelbaum, R. Mapping stellar content to dark matter haloes. II. Halo mass is the main driver of galaxy quenching. *Mon. Not. R. Astron. Soc.* **457**, 4360–4383 (2016).
38. Wang, H. et al. ELUCID. IV. Galaxy quenching and its relation to halo mass, environment and assembly bias. *Astrophys. J.* **852**, 31 (2018).
39. Guo, H., Jones, M.-G., Wang, J. & Lin, L. Star formation and quenching of central galaxies from stacked HI measurements. Preprint at https://arXiv:2105.13505 (2021).
40. Tinker, J. L. Testing galaxy quenching theories with scatter in the stellar-to-halo mass relation. *Mon. Not. R. Astron. Soc.* **467**, 3533–3541 (2017).
41. Mitra, S., Davé, R., Simha, V. & Finlator, K. Equilibrium model prediction for the scatter in the star-forming main sequence. *Mon. Not. R. Astron. Soc.* **464**, 2766–2776 (2017).
42. Tacconi, L. J. et al. High molecular gas fractions in normal massive star-forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).
43. Davé, R., Finlator, K. & Oppenheimer, B. D. An analytic model for the evolution of the stellar, gas and metal content of galaxies. *Mon. Not. R. Astron. Soc.* **421**, 98–107 (2012).
44. Robson, D. & Davé, R. X-ray emission from hot gas in galaxy groups and clusters in SIMBA. *Mon. Not. R. Astron. Soc.* **498**, 3061–3076 (2020).
45. Appleby, S., Davé, R., Sorini, D., Storey-Fisher, K. & Smith, B. The low redshift circumgalactic medium in SIMBA. Preprint at https://arxiv.org/abs/2102.10126 (2021).
46. Davé, R., Finlator, K. & Oppenheimer, B. D. Galaxy evolution in cosmological simulations with outflows. II. Metallicities and gas fractions. *Mon. Not. R. Astron. Soc.* **416**, 1354–1376 (2011).
47. Davies, J. J., Crain, R. A., Oppenheimer, B. D. & Schaye, J. The quenching and morphological evolution of central galaxies is facilitated by the feedback-driven expulsion of circumgalactic gas. *Mon. Not. R. Astron. Soc.* **491**, 4462–4480 (2020).
48. Kennicutt, R. C. & Evans, N. J. Star formation in the Milky Way and nearby galaxies. *Annu. Rev. Astron. Astrophys.* **50**, 531–608 (2012).
49. Davé, R., Thompson, R. & Hopkins, P. F. MUFASA: galaxy formation simulations with meshless hydrodynamics. *Mon. Not. R. Astron. Soc.* **462**, 3265–3284 (2016).
50. Muratov, A. L. et al. Gusty, gaseous flows of FIRE: galactic winds in cosmological simulations with explicit stellar feedback. *Mon. Not. R. Astron. Soc.* **454**, 2691–2713 (2015).
51. Anglés-Alcázar, D. et al. The cosmic baryon cycle and galaxy mass assembly in the FIRE simulations. *Mon. Not. R. Astron. Soc.* **470**, 4698–4719 (2017).
52. Anglés-Alcázar, D. et al. Torque-limited growth of massive black holes in galaxies across cosmic time. *Astrophys. J.* **800**, 127 (2015).
53. Anglés-Alcázar, D., Davé, R., Faucher-Giguère, C.-A., Özel, F. & Hopkins, P. F. Gravitational torque-driven black hole growth and feedback in cosmological simulations. *Mon. Not. R. Astron. Soc.* **464**, 2840–2853 (2017).
54. Choi, E., Ostriker, J. P., Naab, T. & Johansson, P. H. Radiative and momentum-based mechanical active galactic nucleus feedback in a three-dimensional galaxy evolution code. *Astrophys. J.* **754**, 125 (2012).
55. Katsianis, A. et al. The specific star formation rate function at different mass scales and quenching: a comparison between cosmological models and SDSS. *Mon. Not. R. Astron. Soc.* **500**, 2036–2048 (2021).
56. Thomas, N., Davé, R., Anglés-Alcázar, D. & Jarvis, M. Black hole – galaxy correlations in SIMBA. *Mon. Not. R. Astron. Soc.* **487**, 5764–5780 (2019).
57. Li, Q., Narayanan, D. & Davé, R. The dust-to-gas and dust-to-metal ratio in galaxies from $z=0$ to 6. *Mon. Not. R. Astron. Soc.* **490**, 1425–1436 (2019).
58. Appleby, S., Davé, R., Kraljic, K., Anglés-Alcázar, D. & Narayanan, D. The impact of quenching on galaxy profiles in the SIMBA simulation. *Mon. Not. R. Astron. Soc.* **494**, 6053–6071 (2020).

59. Diemer, B., More, S. & Kravtsov, A. V. The pseudo-evolution of halo mass. *Astrophys. J.* **766**, 25 (2013).
60. Conroy, C., Gunn, J. E. & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699**, 486–506 (2009).
61. Conroy, C. & Gunn, J. E. The propagation of uncertainties in stellar population synthesis modeling. III. Model calibration, comparison and evaluation. *Astrophys. J.* **712**, 833–857 (2010).
62. Cui, W. et al. Characterizing diffused stellar light in simulated galaxy clusters. *Mon. Not. R. Astron. Soc.* **437**, 816–830 (2014).
63. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
64. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20 (2007).
65. Millman, K. J. & Aivazis, M. Python for scientists and engineers. *Comput. Sci. Eng.* **13**, 9–12 (2011).

Acknowledgements

We acknowledge helpful discussions with M. van Daalen, J. Matthee, K. Kraljic, D. Sorini, N. Thomas and Y. Zu. We thank R. Thompson for developing Caesar, and the yt team for development and support of yt. W.C. acknowledges the support from the China Manned Space Program through its Space Application System. W.C. and J.A.P. acknowledge support from the European Research Council under grant No. 670193 (the COSFORM project). R.D. acknowledges support from the Wolfson Research Merit Award programme of the UK Royal Society. W.C. and R.D. acknowledge support from the STFC AGP Grant ST/V000594/1. D.A.-A. acknowledges support by NSF grant AST-2009687 and by the Flatiron Institute, which is supported by the Simons Foundation. X.Y. acknowledges support from the National Science Foundation of China (grant Nos. 11833005, 11890692 and 11621303). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility. The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

Author contributions

W.C. conceived the project. R.D. performed the simulation with contributions from D.A.-A. and provided the Caesar catalogue. W.C. designed and performed the analysis. W.C., R.D., J.A.P, D.A.-A. and X.Y. interpreted the results. W.C. wrote the manuscript with contributions from R.D., J.A.P, D.A.-A. and X.Y.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41550-021-01404-1.

Correspondence and requests for materials should be addressed to W.C.

Peer review information *Nature Astronomy* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.