

Jellyfish galaxies with the IllustrisTNG simulations – citizen-science results towards large distances, low-mass hosts, and high redshifts

Elad Zinger^{1,2,★}, Gandhali D. Joshi^{1,2,3}, Annalisa Pillepich^{1,2}, Eric Rohr^{1,2} and Dylan Nelson^{1,4}

¹Centre for Astrophysics and Planetary Science, Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

²Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

³Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

⁴Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany

Accepted 2023 November 28. Received 2023 November 26; in original form 2023 April 18

ABSTRACT

We present the ‘Cosmological Jellyfish’ project – a citizen-science classification program to identify jellyfish (JF) galaxies within the IllustrisTNG cosmological simulations. JF are satellite galaxies that exhibit long trailing gas features – ‘tails’ – extending from their stellar body. Their distinctive morphology arises due to ram-pressure stripping (RPS) as they move through the background gaseous medium. Using the TNG50 and TNG100 simulations, we construct a sample of $\sim 80\,000$ satellite galaxies spanning an unprecedented range of stellar masses, $10^{8.3}–10^{12.3} M_{\odot}$, and host masses, $M_{200,c} = 10^{10.4}–10^{14.6} M_{\odot}$ back to $z = 2$. Based on this sample, $\sim 90\,000$ galaxy images were presented to volunteers in the citizen-science Zooniverse platform, who were asked to determine whether the galaxy image resembles a JF. Based on volunteer votes, each galaxy received a score determining if it is a JF or not. This paper describes the project, the inspected satellite sample, the methodology, and the classification process that resulted in a data set of 5307 visually identified JF galaxies. We find that JF is common in nearly all group- and cluster-sized systems, with the JF fraction increasing with host mass and decreasing with satellite stellar mass. We highlight JF galaxies in three relatively unexplored regimes: low-mass hosts of $M_{200,c} \sim 10^{11.5}–10^{13} M_{\odot}$, radial positions within hosts exceeding the virial radius $R_{200,c}$, and at high redshift up to $z = 2$. The full data set of our JF scores is publicly available and can be used to select and study JF galaxies in the IllustrisTNG simulations.

Key words: galaxies: evolution – galaxies: formation – galaxies: haloes.

1 INTRODUCTION

One of the main determinants of the evolutionary pathway of a galaxy is the environment it resides in. Observations have long shown stark differences in a range of properties between galaxies that reside in the centres of their host haloes versus galaxies that are satellites at larger distances (Dressler 1980). In particular, satellite galaxies are more likely to be quenched compared to field and central galaxies of the same stellar mass (Lewis et al. 2002; Gómez et al. 2003; Peng et al. 2010), and have lower gas fractions, including both atomic and molecular components (Gavazzi et al. 2005; Fumagalli et al. 2009; Catinella et al. 2013; Boselli et al. 2014).

A set of physical processes, that are environmental as opposed to secular, lead to these differences, primarily through the removal of gas from the galaxy body (Cortese, Catinella & Smith 2021). Some of these processes affect all components of the galaxy via gravitational interactions – such as tidal stripping (Gnedin 2003a; Villalobos, De Lucia & Murante 2014) or galaxy harassment (Moore et al. 1996; Gnedin 2003b). Others arise from hydrodynamical effects, including ram-pressure stripping (RPS) (Gunn & Gott 1972) and viscous stripping, Nulsen (1982), or thermo-dynamical channels (e.g. thermal evaporation, Cowie & Songaila 1977), and so affect only

the gaseous component, leaving the stellar structure of the galaxy unchanged.

RPS occurs when the ram pressure is generated by the motion of a satellite through the ambient medium. The ram-pressure force is $\propto \rho_{\text{medium}} v_{\text{sat}}^2$, where ρ_{medium} is the gas density of the medium and v_{sat} is the velocity of the satellite with respect to the medium. Stripping occurs when this force exceeds the gravitational restoring force of the galaxy (see Boselli, Fossati & Sun 2022, for a recent review). Viscous (or turbulent) stripping removes gas from the outer layers of a galaxy through momentum transfer by viscosity. This process is thought to be less important for satellite galaxies than RPS (Roediger & Brüggen 2008; Roediger 2009). Both processes lead to extended gas structures that emanate from the stellar body and trail behind it, a configuration that resembles a jellyfish (JF). Due to this similarity, such galaxies have come to be known as ‘JF’ galaxies and have by now been observed across a broad wavelength range (e.g. Bekki 2009; Chung et al. 2009; Smith et al. 2010; Ebeling, Stephenson & Edge 2014; Poggianti et al. 2017).

Due to its dependence on the medium properties, RPS is expected to be more important in more massive haloes and in the inner regions of the haloes since the ambient medium is denser and the infall velocities are higher, due to the deeper potential wells. In addition, smaller galaxies, with weaker gravitational binding, are expected to be more affected by RPS. These trends are by and large consistent with observations (Roberts et al. 2021b; Boselli, Fossati & Sun 2022).

* E-mail: elad.zinger@mail.huji.ac.il

For these reasons, observational efforts to find JF galaxies usually focus on the inner regions of galaxy cluster-sized systems, i.e. in hosts with $M_{200,c} \gtrsim 10^{14} M_{\odot}$ (though some surveys have detected JF galaxies in group-sized hosts, e.g. Poggianti et al. 2016; Vulcani et al. 2018; Roberts et al. 2021a; Cortese, Catinella & Smith 2021) and at low redshifts of $z < 1$ (e.g. Boselli et al. 2019), where such massive systems are more common. However, some theoretical studies have suggested that RPS can occur at and even beyond the virial radius (Bahé et al. 2013; Cen 2014; Jaffé et al. 2015; Zinger et al. 2018; Ayromlou et al. 2019), and JF galaxies have also been observed to occupy these regions (Chung et al. 2009; Boselli et al. 2018; Gullieuszik et al. 2020; Roberts et al. 2021b).

Galaxies affected by ram pressure have been detected first in the radio (Miley et al. 1972; Gavazzi & Perola 1978), with subsequent multiwavelength observations revealing that the tails of JF galaxies are multiphase and can contain ionized gas (H α : Boselli et al. 2016; Gavazzi et al. 2018), neutral Hydrogen (HI: Shostak et al. 1982; Kenney, Gorkom & Vollmer 2004; Healy et al. 2021), molecular gas (Jáchym et al. 2014, 2017; Verdugo et al. 2015), and hot-ionized gas (X-rays: Machacek et al. 2006; Sun et al. 2006; Wood et al. 2017).

Many of the observational efforts have focused on single galaxies or the population of a single host cluster (e.g. Boselli et al. 2018), but recent years have seen the advent of several surveys targeting larger samples of JF galaxies, such as the GAs Stripping Phenomena in galaxies project (GASP, Poggianti et al. 2017), LOFAR Two-Meter Sky Survey (LoTSS: Roberts et al. 2021b), and OSIRIS Mapping of Emission-line Galaxies (OMEGA: Roman-Oliveira et al. 2019), each of which compiled JF samples of 50–70 objects. These surveys have made statistical studies of ~ 100 observed JF galaxies possible (Smith et al. 2022; Peluso et al. 2022).

Numerical simulations enable us to build a theoretical picture of JF galaxies. Idealized ‘wind-tunnel’ type simulations focus on a single galaxy in high spatial and temporal resolution and can explore the impact of different physical effects (Roediger & Brüggén 2007; Roediger 2009; Tonnesen & Bryan 2009). Large-scale cosmological simulations such as Illustris (Vogelsberger et al. 2014; Genel et al. 2014; Vogelsberger et al. 2014; Sijacki et al. 2015), EAGLE (Crain et al. 2015; Schaye et al. 2015), and IllustrisTNG (Pillepich et al. 2018a, and references below), have achieved sufficiently high resolution, large-enough volumes, and physical sophistication to generate large samples of JF galaxies. These large-volume simulations enable statistical studies over a large range of satellite and host properties, and the ability to follow the history of each JF galaxy in a cosmological context (Yun et al. 2019). Zoom-in simulations of galaxy cluster-sized systems, for example, RomulusC (Tremmel et al. 2019), offer additional opportunities to study the evolution JF galaxies in the high-resolution, cosmological setting (Ricarte et al. 2020). Semi-analytic models also commonly incorporate RPS effects in order to successfully model satellite galaxy evolution in dense environments (Somerville et al. 2008; Lagos et al. 2018; Zinger et al. 2018; Ayromlou et al. 2019).

Despite the advantages of numerical studies, it is not computationally possible to include all the relevant physical processes in simulations. For example, many simulations do not include magnetic fields although they have been suggested to be important in shaping the tails of JF galaxies (Ruszkowski et al. 2014; Tonnesen & Stone 2014; Müller et al. 2021). Large-scale cosmological simulations, such as IllustrisTNG, do not typically include physical viscosity and thus do not model gas stripping by viscous momentum transfer. For this reason, JF galaxies in cosmological galaxy simulations, which are the subject of this study, are shaped predominantly by RPS.

To study JF galaxies within a large sample of systems, whether observed or simulated, one must first find them. The asymmetric tail pattern that is the hallmark of these objects is usually easy to identify visually. And indeed, in most observational endeavours (e.g. Ebeling, Stephenson & Edge 2014; Poggianti et al. 2017; Roman-Oliveira et al. 2019; Roberts et al. 2021b), the classification of galaxies as JF relies on a visual inspection by the observer(s) and, even when automated identification methods are used, the results are usually visually confirmed (McPartland et al. 2016; Roberts et al. 2021a).

In an earlier project (Yun et al. 2019, hereafter Yun19), our team identified and studied a large sample of JF galaxies in the TNG100 simulation, one of the flagship runs of the IllustrisTNG project. To find JF galaxies among the many thousands of simulated satellites, five members of the research team carried out a visual classification of galaxy images depicting gas column density. The project included ~ 2600 inspected satellites selected from four snapshots of the TNG100 simulation box, ranging between $z = 0.6$ and 0, of which about 800 were identified as JF galaxies.

In Yun19, we used this sample to study the demographics of JF galaxies, finding that their frequency increases with host mass and decreases with satellite stellar mass. No strong dependence was detected across the redshift range. Roughly equal numbers of JF galaxies were found on both infalling and outgoing trajectories. We also found that JF galaxies exhibit higher velocities, higher Mach numbers and experience stronger ram-pressure than other satellites, showing that the physical driver of the tails in JF galaxies is indeed ram-pressure.

In this work, we build upon and extend the study initiated in Yun19. Namely, we use the TNG50 and TNG100 simulations of the IllustrisTNG project to generate an unprecedentedly large sample of about 90 000 images (of more than 80 000 galaxies) in which we search for JF galaxies. The sample spans satellite stellar masses of $10^{8.3} - 10^{11.5} M_{\odot}$, host masses of $10^{10.5} - 10^{14.5} M_{\odot}$ and extends back to $z = 2$, exceeding currently existing observational and simulation-based studies. Furthermore, the IllustrisTNG simulations are fully cosmological, include magnetohydrodynamics and hence magnetic fields, as well feedback models from stars as well as super massive black holes. To identify JF within the gas-based images, we developed a citizen-science project: the Cosmological Jellyfish Zooniverse Project (hereafter CJF Zooniverse project). Its goal was to produce reliable identifications of JF galaxies for our $\sim 90\,000$ images generated from IllustrisTNG. Our citizen-science project was launched on 2021 June 14 and was completed in two consecutive phases: in 2021 June (Phase 1) and between 2021 August and November (Phase 2), involving a total of more than 6000 volunteers. This paper collects and analyses the classifications produced therein, presents the first scientific results based upon them, and serves as a reference guide for future scientific and outreach usages of the data i.e. of the classifications, which we publicly release here.

Citizen science is a relatively recent development – one of the first, large-scale citizen-science projects, Galaxy Zoo (Lintott et al. 2008), was founded to contend with the unprecedented large data sets from the Sloan Digital Sky Survey and the need to identify the morphology of nearly a million galaxies. Over 100 000 volunteers took part in the various classification tasks. Galaxy Zoo has been active since its launch in 2007 to this day. Volunteers have even discovered several unknown objects and novel phenomena (e.g. Cardamone et al. 2009; Lintott et al. 2009). After these initial successes the Zooniverse platform,¹ for citizen-science projects was established

¹www.zooniverse.org

(Borne et al. 2009; Borne & Zooniverse Team 2011; Smith et al. 2011). The platform enables research scientists in a variety of fields to enlist the help of volunteers in carrying out scientific tasks.

For our work, the citizen-science approach to classify JF galaxies was a natural one. Their unique characteristics, i.e. asymmetrical gas tails, make them easy to identify visually, even for participants with no prior knowledge. Further, inspecting tens of thousands of galaxy images is a formidable challenge, since their sheer number (over 13 times more than in Yun19) exceeds our team’s ability to inspect them all. While automated identification of tails and JF morphologies will be an important tool in the future, a citizen-science project provided a compelling way to complete the classification task. Overall, the existence of the Zooniverse framework, access to a pre-existing pool of interested volunteers, and a high level of engagement and dedication all made the process highly successful.

In this paper, we describe our CJF Zooniverse project in detail, and present results on the demographics of JF galaxies according to the outcome of the IllustrisTNG simulations. Section 2 describes the data set we use, i.e. galaxy images generated from the TNG50 and TNG100 simulations, and details the classification process carried out in the Zooniverse platform. In Section 3, we describe how the volunteer classification data are synthesized to generate a galaxy ‘score’ and compare these scores to expert classifications, validating the citizen-science results. We then set a threshold score for JF galaxy identification in Section 4 and suggest a statistical framework for evaluating the confidence level associated with a given threshold score. With a fully classified sample in hand, we present results about the demographics of the JF galaxy population, and their host haloes in Section 5, with an emphasis on the JF population found in low-mass hosts and at large radial distances from the host centre. In Section 6, we discuss the validity of our classification method, and also the public-outreach value of our project. Finally, we summarize our findings in Section 7.

2 THE CJF ZOONIVERSE PROJECT

2.1 The TNG50 and TNG100 cosmological galaxy simulations

Throughout this paper and to construct the sample of satellite galaxies to be visually inspected, we make use of two simulations from the IllustrisTNG project,² a suite of magnetohydrodynamic cosmological simulations carried out in three volumes of varying size and resolution. Namely, we use the flagship runs called TNG100 (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) and TNG50 (Pillepich et al. 2019; Nelson et al. 2019b). All IllustrisTNG simulations were run with the AREPO code (Springel 2010; Pakmor & Springel 2013) and are based on a physical model of galaxy formation (the IllustrisTNG model, Weinberger et al. 2017; Pillepich et al. 2018a) that has been shown to reproduce, with a reasonable level of accuracy, a large number of observational properties of galaxies.

TNG50 and TNG100 evolve cubic volumes of comoving side length of roughly 50 and 100 Mpc, respectively. By combining both these simulations, we obtain a large statistical sample of satellite galaxies residing in a large variety of hosts – from Milky Way-sized haloes through groups and clusters of galaxies (up to $10^{14.6} M_\odot$ in total mass). The higher resolution TNG50 box allows us to probe the lower mass satellites ($\gtrsim 10^{8.3} M_\odot$) regime, as well as providing a more detailed outcome. Galaxies are resolved with baryonic mass

resolution of 8.5×10^4 and $1.4 \times 10^6 M_\odot$ in TNG50 and TNG100, respectively. The full simulation data of TNG50 and TNG100 are publicly available (see Nelson et al. 2019a, for details).

The cosmological framework for these simulations is a Lambda cold dark matter cosmological model with the parameter values based on the Planck Collaboration XIII (2016) data: cosmological constant $\Omega_\Lambda = 0.6911$, matter density $\Omega_m = \Omega_{dm} + \Omega_b = 0.3089$, with a baryonic density of $\Omega_b = 0.0486$, Hubble parameter $h = 0.6774$, normalization $\sigma_8 = 0.8159$, and spectral index $n_s = 0.9667$.

2.2 Selection of TNG50 and TNG100 satellite galaxies inspected in the CJF project

Among the many tens of thousands of satellite galaxies simulated within TNG50 and TNG100 across cosmic epochs, a sample (denoted from now on as ‘inspected’) was selected for visual inspection from the SUBFIND subhalo catalogues at various snapshots. The selection was based on the galaxy stellar mass and gas fraction, as well as central or satellite status within their host Friends-of-Friends (FoF) halo. Unless otherwise specified, all stellar and gas masses are measured within a radius r_{gal} , which is set to be twice the radius that encloses half of the stellar mass in a given subhalo ($r_{gal} \equiv 2r_{1/2, stars}$). Furthermore, we exclude subhaloes that are likely to be clumps of matter and unlikely to be galaxies of cosmological origin, based on the SubhaloFlag defined in Nelson et al. (2019a). In particular, we applied the following criteria to select our galaxy sample:

- (i) Only satellites, i.e. excluding central galaxies.
- (ii) Satellite stellar mass: $M_* > 10^{8.3} M_\odot$ in the case of TNG50, and $M_* > 10^{9.5} M_\odot$ in the case of TNG100.
- (iii) Gas fraction: $f_{gas} > 0.01$, with f_{gas} defined either as $M_{gas, r < r_{gal}} / M_{*, r < r_{gal}}$ (Phase 1), or $M_{gas, all} / M_{*, r < r_{gal}}$ (Phase 2).

In Phase 1, galaxies were selected from the following snapshots:

- (i) TNG50 and TNG100: Full snapshots up to redshift $z = 2$ i.e. snapshots 99, 91, 84, 78, 72, 67, 59, 50, 40, and 33. The time interval between pairs of these snapshots is ~ 1 Gyr.

While in Phase 2 the snapshot selection was as follows:

- (i) TNG50 and TNG100: Full snapshots up to redshift $z = 2$, excluding galaxies already considered in Phase 1. These galaxies are those with $f_{gas, 2r_{1/2, *}} < 0.01$ but $f_{gas, all} > 0.01$, making the samples complete down to the lower of the two gas fraction criteria.
- (ii) TNG50: All snapshots up to $z = 0.5$, i.e. snapshots 68–98 (inclusive), excluding those already considered in Phase 1. The time interval between the snapshots is ~ 150 Myr.

Furthermore, in Phase 2 we also included, as a test, an additional 2918 images from TNG50 and 5844 images from TNG100 of the same galaxies selected as above solely from snapshots 99 and 67 (i.e. at $z = 0$ and 0.5) and shown in a preferred, rather than random, orientation: these will be described and studied in Section 6.3.

Note that no selection was imposed a priori on the mass of the underlying dark-matter host haloes. Furthermore, 20 galaxies were removed from the Phase 1 inspected sample, as their images were deemed to be too messy for classification. These galaxies were inspected by the expert team members and manually assigned a score of 0, as described in later sections. Importantly, the galaxies selected for inspection are biased towards larger gas fractions than the randomly selected satellite in groups and clusters at any given time and stellar mass: this is because they need to have at least some gas, otherwise they could not exhibit tails of stripped gas and could not be identified as JF.

²www.tng-project.org

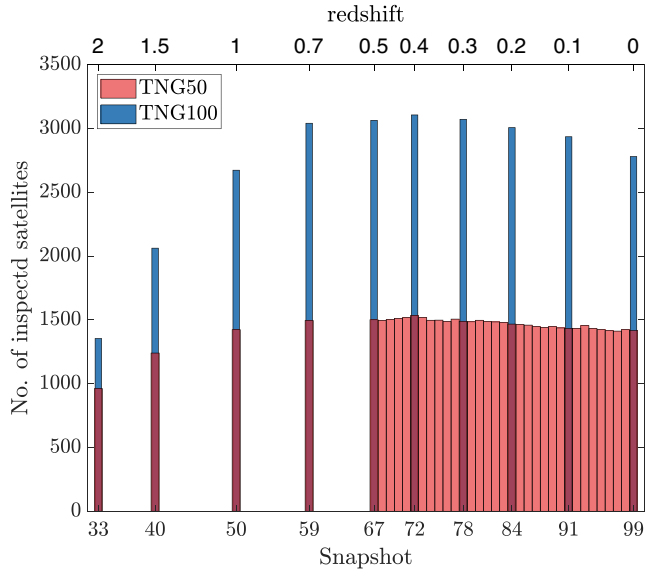


Figure 1. Number of satellite galaxies selected for the visual inspection in our CJF Zooniverse project from the TNG50 (red) and TNG100 (blue) simulations, at each selected snapshot. The corresponding redshift values are shown in the upper axis. We deliberately chose to inspect TNG50 satellites at $z \leq 0.5$ as frequently as allowed by the available output data, i.e. approximately every 150 Myr.

Table 1. Number of hosts and satellites from the TNG50 and TNG100 simulations proposed for visual inspection in our CJF Zooniverse project, in three redshift bins. In total, 80 704 satellites were classified based on randomly oriented images.

Redshift	TNG50		TNG100	
	Hosts	Satellites	Hosts	Satellites
$0 \leq z \leq 0.5$	17 085	48 491	6668	17 964
$0.5 < z \leq 1$	1146	2918	2460	5714
$1 < z \leq 2$	1070	2201	1740	3416
Total	19 301	53 610	10 868	27 094

In total, we selected for the visual inspection on Zooniverse 13 592 and 24 394 galaxies from TNG50 and TNG100, respectively, in Phase 1 and a further 40 001 and 2697 galaxies in Phase 2, respectively. The total number of galaxies inspected during the two phases of our CJF project is 80 704 (including the 20 ‘messy’ galaxies), and they reside within 30 169 hosts. We compared these numbers to the total number of satellite galaxies in the TNG50 and TNG100 simulations, above the respective mass thresholds of $10^{8.3}$ and $10^{9.5} M_{\odot}$, and their respective hosts. We find that the CJF-inspected sample comprises ~ 90 per cent of all relevant satellites at high redshifts ($z = 2$), with this fraction dropping at low redshifts to ~ 60 per cent for TNG100 and ~ 50 per cent for TNG50. The host haloes in the CJF sample comprise between 96 (low redshifts) and 99 per cent at $z = 2$ of all relevant hosts.

The total number of images put up for inspection was 89 446, which includes the additional 8762 images of galaxies repeated from the preferred viewpoint. In all that follows, unless specified otherwise, the tables, figures, and results, all describe the inspected sample of 80 704 galaxies, with the identification of these galaxies as JF or not relying on images generated from a random viewing angle.

In Fig. 1, we show the number of objects in each snapshot/redshift from the TNG50 and TNG100 simulations. In Table 1, we list the

number of inspected objects, in particular, of hosts and satellites in the TNG50 and TNG100 samples divided into three relevant redshift ranges. Since our total sample of inspected galaxies includes all the snapshots of TNG50 at $z \leq 0.5$, we can see that TNG50 objects from this time-frame account for ~ 58 per cent of the satellites and ~ 55 per cent of the hosts, even though TNG50 simulates a much smaller volume than TNG100.

In Fig. 2, we show the basic demographics of the TNG50 and TNG100 selected and thus inspected objects, over all selected redshifts. In Fig. 2(a), we show the distribution of the satellite stellar masses of the galaxies in the inspected sample, whereby the imposed lower mass limits are clearly evident. The high stellar mass end is shown in detail in the inset: we can study satellites up to a few $10^{11} M_{\odot}$ in stars. Even though the TNG100 volume is roughly eight times larger, there are more galaxies in the TNG50 inspected sample, mainly due to there being more snapshots included from the TNG50 simulation (see Fig. 1).

Fig. 2(b) quantifies the mass distribution of the hosts in which the galaxies reside at the time of inspection. The host mass shown is $M_{200,c}$ of the host), i.e. the mass enclosed within a radius $R_{200,c}$, which, in turn, is defined as the radius where the mean density of the halo is equal to 200 times the critical density of the universe at that time. The TNG50 inspected sample allows us to probe low-mass satellite galaxies and, by extension, low-mass hosts. The TNG100 inspected sample supplies us with a very large number of group- and cluster-sized hosts. In Fig. 2(c), we show the distribution of the satellite-to-host mass ratio, where again we see that TNG50 allows us to explore a wide range of satellite–host interactions.

Finally, the total number of satellites found and inspected within all hosts in a given mass range is shown in Fig. 2(d). Due to the volume of TNG100, there are many satellites in high-mass hosts. Overall, the 80 704 satellite galaxies in the CJF Zooniverse encompass a large range of satellite stellar masses, host halo masses and redshift, namely: $10^{8.3} - 10^{12.3} M_{\odot}$, $M_{200,c} = 10^{10.4} - 10^{14.6} M_{\odot}$, and $z \leq z \leq 2$, respectively. As we expand upon in section 6.2 and in Rohr et al. (2023), it is important to note that not all satellites in the CJF project are unique, in that many of them represent different evolutionary stages of the same galaxy selected and inspected at different redshifts.

In Fig. 3, we show how many inspected satellites are found in the individual hosts at any given mass. The distribution of hosts is shown by a colour-map that signifies the number of inspected hosts within a given host mass versus satellite number bin. The *total* number of hosts inspected throughout the project is given in the top-left corner (see also Table 1). The distributions from TNG50 and TNG100 are shown separately in Figs 3(a) and (b), with a breakdown of the host population into three redshift bins shown in the three smaller panels. As expected, the number of hosts, and the number of satellites within them, grows towards lower redshift.

Because the galaxy and halo populations from large-volume cosmological simulations like IllustrisTNG are volume limited by construction, the majority of hosts are found in the low-mass range, with mass $M_{200,c} \lesssim 10^{13} M_{\odot}$ (as already manifest in Fig. 2): In such hosts, the typical number of inspected satellites is of order of a few and very rarely above 10. The higher resolution and lower mass limit afforded by TNG50 result in slightly higher numbers of satellites per host. The larger volume of TNG100 results in many more hosts at masses of $M_{200,c} \gtrsim 10^{13} M_{\odot}$.

2.3 Visual classification and identification of JF

As discussed above, the identification of JF galaxies was carried out through visual inspection on the Zooniverse platform for citizen

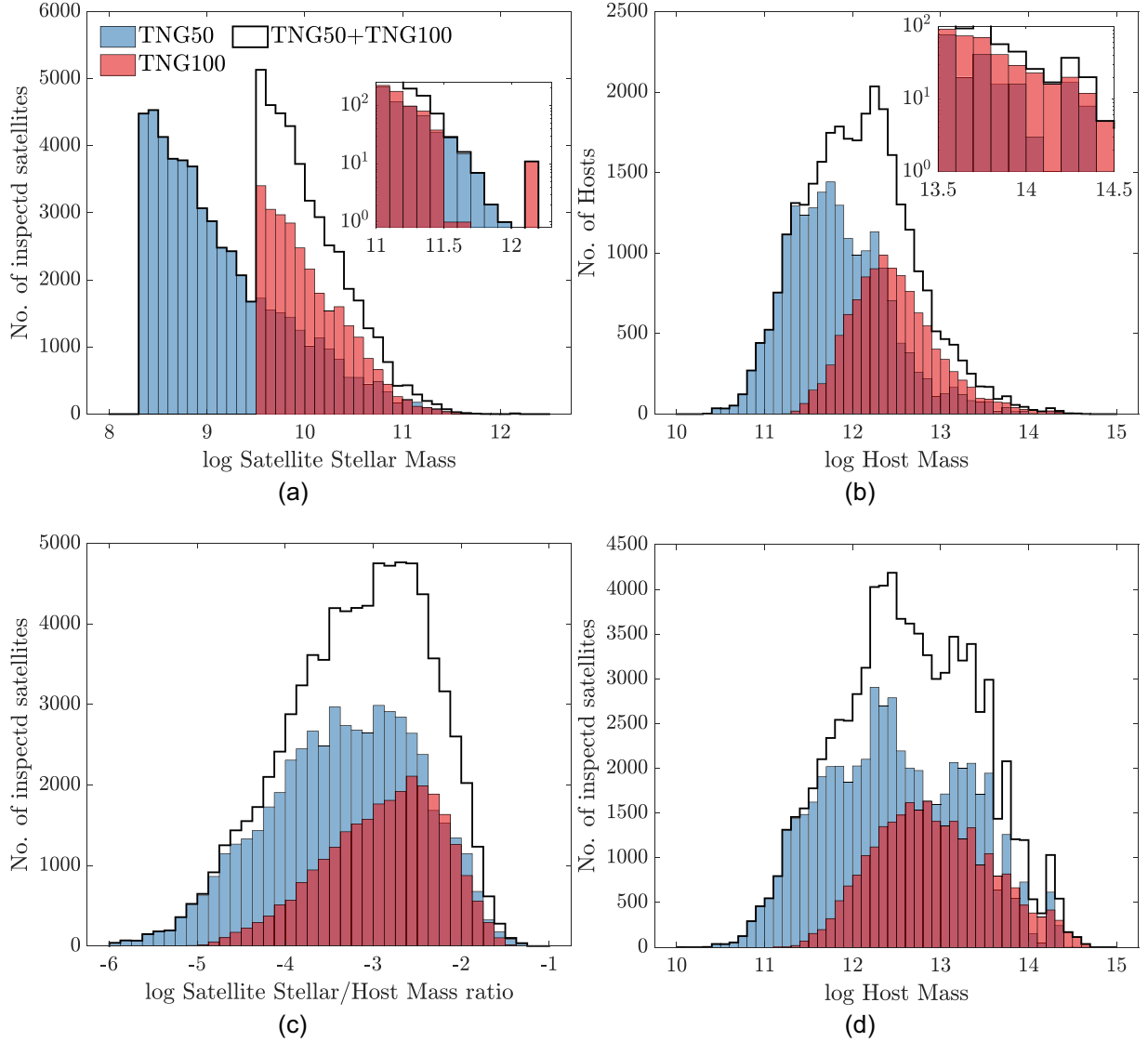


Figure 2. Demographics of the satellite galaxies selected for visual inspection in our CJF Zooniverse project. The panels show the distribution of satellite stellar mass (a), host mass $M_{200,c}$ (b), the satellite stellar to host mass ratio (c) and the total number of satellites in hosts of a given mass (d) across all selected snapshots ($0 \leq z \leq 2$, see the text for details). The distributions for the satellites selected from the TNG50 and TNG100 simulations, and thus visually inspected, are shown separately in blue and red histograms, respectively, with the black line showing the combined population from both simulation. In panels (a) and (b), the inset shows the distribution at the high-mass end.

science.³ The CJF project considered images of all the galaxies in the sample defined in Section 2.2. As we describe below, the image format as well as the classification process are geared to discover JF galaxies in which the gas tails originate from the galaxy ISM. A total of 6494 volunteers participated in the classification effort, with a total of 1.8 million classifications. Here, we describe the proposed tasks, the workflow, and the characteristics of the inspected images.

In our CJF Zooniverse project, after a short training process, volunteers were shown one galaxy image at a time and were asked to classify the galaxy in the centre of the image by answering a simple yes/no question: ‘Do you think that the galaxy at the centre looks like a JF?’. Once an image was classified by 20 different people, it was retired from the image pool. For context, the original Galaxy Zoo project relied on ~ 38 inspectors per image (Lintott et al. 2008).

As detailed below, the training included a number of examples and guidelines, under the form of a tutorial and a field guide, available to the volunteers at any step of the classification.

2.3.1 Image specifications

The images generated for the classifications consist of maps of two main matter components: (i) gas mass surface density shown as a coloured 2D histogram and (ii) stellar mass surface density shown as white contours overlaid on the gas map. For these, we followed the methodology and visualization technique of Nelson et al. (2019a).

The map of the gas distribution was generated by measuring the (log of the) mass projected on an 800×800 pixel grid of side length $40 \times r_{1/2,*}$ centred on the given galaxy. All of the gas belonging to the host halo was included, but restricted to a cube of side length $40 \times r_{1/2,*}$. The colour range corresponded to $10^5 - 10^8 M_{\odot} \text{ kpc}^{-2}$ in

³www.zooniverse.org/projects/apillepich/cosmological-jellyfish

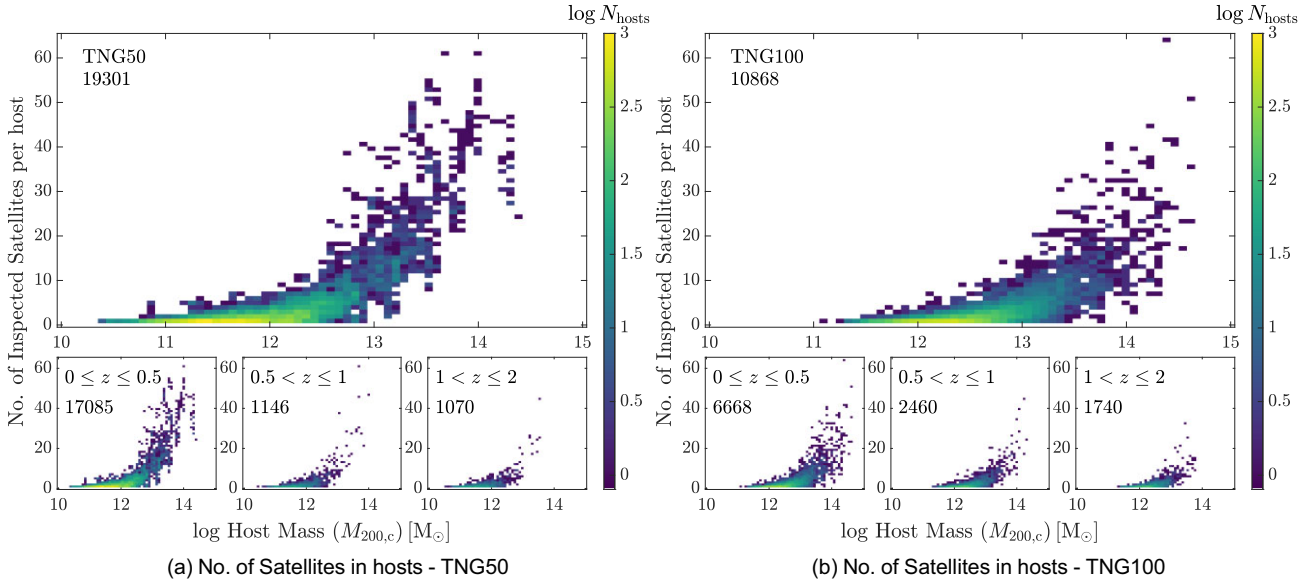


Figure 3. The number of inspected satellites in each individual host, as a function of host mass. The colour-map corresponds to the number of hosts in each bin. The host population of TNG50 (TNG100) is shown in panel (a) and (b). The number of simulated hosts inspected within our CJF Zooniverse project is shown in the top-left corner of each panel. The three smaller panels on the bottom show the inspected host sample in three redshift bins.

column density (colour bars not shown to classifiers). This density range highlights the distribution of gas originating from the satellites’ ISM, and is not conducive for identification of filamentary features that may originate solely from the CGM.

Stellar contours were then overlaid on the gas maps to allow the classifier to determine the extent of the stellar mass in the galaxy under consideration as well as to indicate the presence of other galaxies in the image. In order to generate these contours, we selected all galaxies within the field of view down to a mass limit 0.5 dex lower than that used for the actual inspected sample i.e. $10^{7.8} M_{\odot}$ for TNG50 and $10^9 M_{\odot}$ for TNG100. This lower limit was used to ensure that we captured most galaxies that would be evident in the images. As with the gas maps, we first obtained a 2D histogram of the (log) stellar mass – all gravitationally bound stellar mass to each galaxy – projected on the same grid as that used for the gas maps, but *separately for each galaxy in the field of view*. Contours were then generated based on these 2D histograms, corresponding to 75, 80, and 85 percent of the log of the peak mass surface density for the given subhalo. These levels were determined manually and after inspection of many systems in order to adequately capture the extent of the stellar mass distribution of each galaxy. Finally, the images were saved with a resolution of 300 dpi.⁴ Examples of the images as they appeared on the website can be seen in Figs 4 and 10 and in Appendix B.

The simulated galaxies were projected along random orientations, i.e. from an arbitrary view point irrespective of, for example, the orientation of the stellar disc or the galaxy location within the host halo. To do so, we projected the mass distribution of each galaxy along the z -axis of the simulated volume. A subset of galaxies ($z = 0, 0.5$ snapshots from TNG50 and TNG100) were also projected along an orientation optimized for tail identification, for the purpose

of studying the effect of viewing angle on classification. We discuss this study in Section 6.3.

2.3.2 Visual classification process

The classification process on the Zooniverse platform was open to anyone, with no requirements on previous training or experience. We provided ample background information, as well as an extensive training guide. New volunteers were shown this training guide upon their first entry to the site. In the training guide, a JF galaxy was described as a ‘galaxy (that) exhibits one or more “tails” of gas that stem from the main galaxy body and stretch in one preferred direction. Such galaxies almost look like the JF in the sea!’. The volunteers were asked to focus on the gas features, shown in a density colour map (see Section 2.3.1), in relation to the stellar body of the galaxy, shown by thick white contours (see e.g. Fig. 4). Requiring that the tail be (apparently) attached to the stellar body biases towards identifying tails comprised of ISM gas.

Several examples of what we considered to be clear-cut JF galaxies were shown, but most of the training guide was to point out, via examples, what a JF galaxy is *not*: galaxies with no gas tails, galaxies with very little gas, galaxies with gas tails pointing in many different i.e. opposite directions, galaxies with tails that do not appear to connect to the galaxy and galaxies with very messy gaseous surroundings. In addition, volunteers were asked to treat galaxies with nearby neighbours as non-JF, *even if they exhibited gas tails*, to avoid confusion with merger events.

In addition, we added a training set of images that were previously classified by our research team, either as a part of the Yun19 project or as part of a pilot project that was carried out within the team prior to the public release of the official CJF Zooniverse project (see Section 3.2 for more details). Volunteers were shown images from the training set sporadically, and asked to classify them without being aware that they are from the training set: upon completion of the associated task, they were then notified whether or not their yes/no choice matched the expert one, receiving an immediate feedback on their classification.

⁴For a small number of cases, where the image file size exceeded the platform limit the image was saved at a resolution of 100 dpi. We confirmed that this made no visual difference.

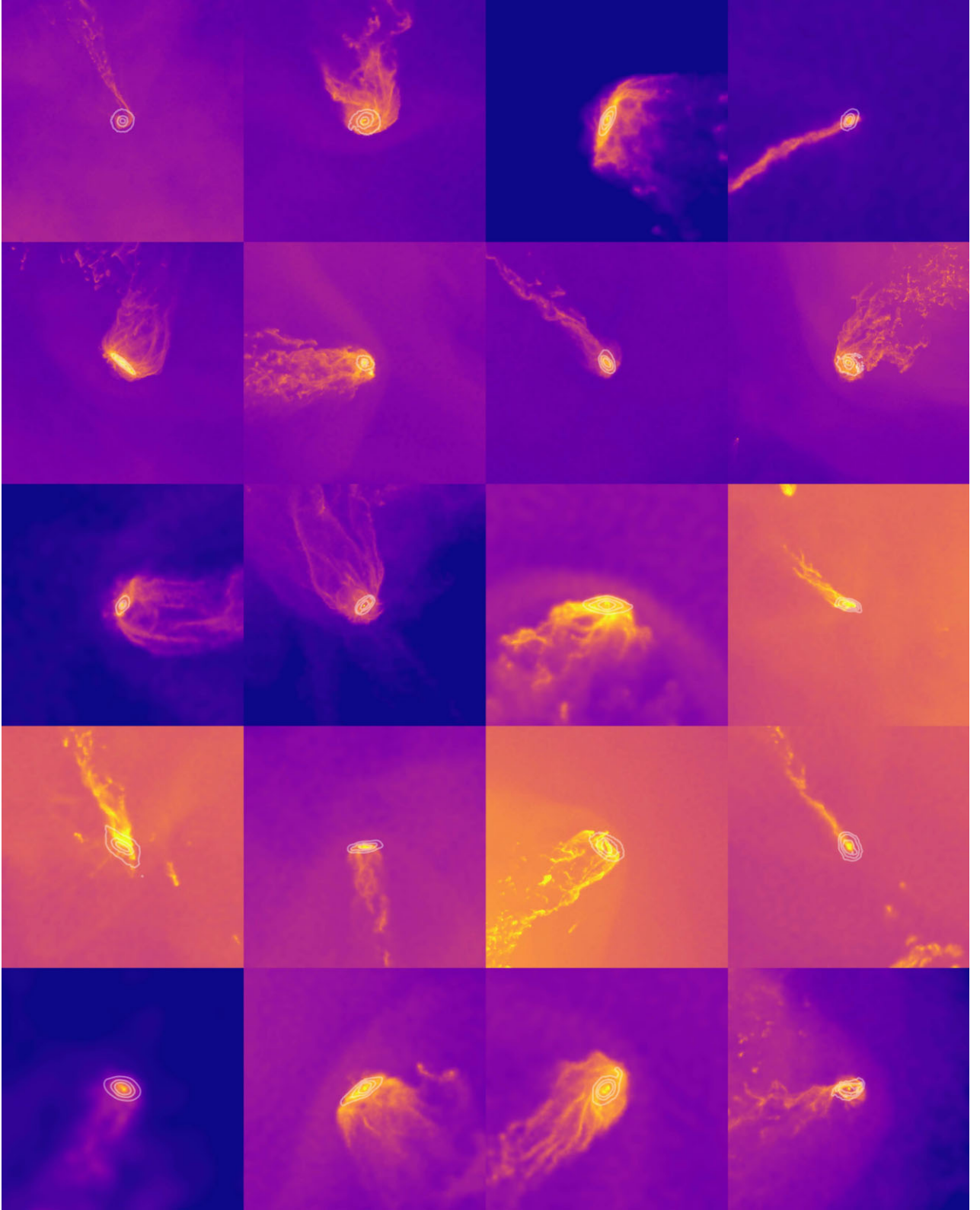


Figure 4. Examples of JF galaxies identified via the CJF Zooniverse project, including galaxies from both the TNG50 and TNG100 simulations at various redshifts. These are the exact stamps shown to the volunteer inspectors, depicting gas column density as colour map and stellar mass density as white transparent contours. The gas density ranges between 10^5 (blue) and 10^8 (yellow) $M_{\odot} \text{ kpc}^{-2}$, the stamps have a side length of $40 \times$ the stellar half-mass radius of the galaxy. All JF galaxies identified in this paper can be seen at www.tng-project.org/explore/gallery/zinger23/. The procedure we use to identify JF galaxies based on the visual classifications is described in Sections 3 and 4.

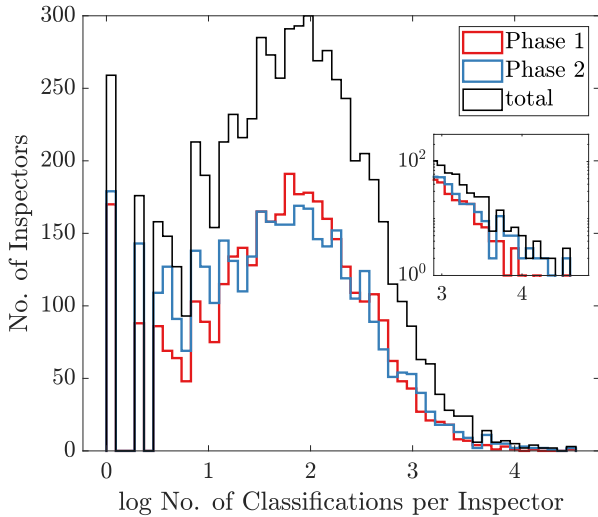


Figure 5. Distribution of the number of inspectors who classified a certain number of images in our CJF Zooniverse project, for the two phases (red and blue lines, respectively) and for the entire inspected sample (black line).

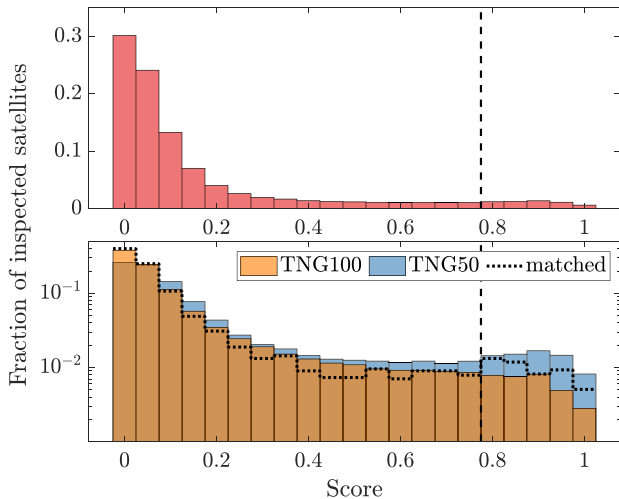


Figure 6. Distribution of the raw JF scores from our CJF Zooniverse project after about 1.8 million classifications. We show results for the entire inspected sample in the top panel, and separately for the TNG50 (blue) and TNG100 (orange) simulated inspected galaxies in the bottom panel. The dotted histogram shows the distribution for a sub-sample of the TNG50 galaxies chosen to mimic the TNG100 sample in terms of mass threshold and snapshots. The y-axis denotes the fraction of the sample population – note that the bottom panel is shown in log scale. The vertical dashed line denotes the threshold chosen in this work (0.8 in a range from 0 to 1) above which we define a galaxy to be a JF.

Finally, volunteers could discuss specific cases on a public forum, and often sought (and received) assistance in the classification process from the members of the research team.

In Fig. 5, we show the distributions of classifications per inspector, quantifying how many volunteers classified a certain number of images, for the two project phases separately as well as for the entire inspected sample. As described above, the classification process was carried out in two phases: Phase 1 included 37 986 galaxies, produced 759 720 classifications (20 classifications per image), while Phase 2 included 51 460 images and resulted in 1029 200 classifications. In summary:

- (i) 6494 inspectors performed a total of 1795 292 classifications – a few objects actually have more than 20 classifications;
- (ii) 4 per cent of the inspectors only classified a single object (of them, 60 per cent were not logged on);
- (iii) 19 per cent of the inspectors classified fewer than 10 objects;
- (iv) the two phases exhibit very similar distributions of participation;
- (v) the average number of classifications per inspector is 276 overall; the median number of classifications per inspector is 45 overall;
- (vi) 19 per cent of all classifications came from anonymous users who did not register on the Zooniverse website or registered volunteers who did not log-on before beginning to classify; and
- (vii) 5 per cent among the inspectors *each* classified more than 1000 images, being responsible for 56 per cent of all classifications;
- (viii) of these, 44 inspectors, each classified more than 5000 objects, being responsible for about 27 per cent of all classifications.

Whereas most inspectors classified several tens or even a few hundred objects each, a small number of very dedicated inspectors (about 5 per cent) are responsible for more than half of all classifications, while nearly a fifth of all inspectors can be considered inexperienced due to them having classified only a handful of images. In light of this, we assess the quality of individual inspectors and weigh the scores accordingly, as we detail in Section 3.3.

3 ASSESSMENT OF THE OUTCOME OF THE CJF ZOONIVERSE CLASSIFICATION

3.1 Raw JF scores

At the end of the classification process, all the classifications of a given object were tallied and a final score between 0 and 20 was assigned to each galaxy. A small percentage of objects (~ 2 per cent in phase 1 and ~ 1.16 per cent in phase 2) received more than 20 classifications due to technical issues. In order to generate a standardized score between 0 and 20 for each of these objects, we created 200 random sub-samples of 20 votes (out of the entire classification pool for that object) and assigned the median value over these 200 scores as the final score for that object. The scores are then normalized to values between 0 and 1.

In Fig. 6, we show the distribution of these raw scores of our inspected galaxy sample, both for the entire inspected sample and also for the TNG50 and TNG100 samples separately. The dashed vertical lines denote the score threshold we choose in this study to define JF galaxies (see also Section 4.2): galaxies with a raw score of 0.8 or higher are dubbed JF or, in other words, 16 of 20 inspectors deemed a given galaxy to be a JF.

By comparing the results from the TNG50 and TNG100 simulations, we find that the TNG50 sample is skewed towards higher scores. This may be due to the higher resolution of the TNG50 simulation run: The features that lead to a classification of JF are more pronounced at higher resolution. However, differences in the two populations exist and may be responsible for the different results without being directly connected to the underlying numerical resolution: The TNG50 sample is dominated by lower mass and lower redshift objects (Table 1), which are more likely to be JF galaxies (see e.g. Yun19 and the next sections).

To shed light on the reasons of the differences between the TNG50 and TNG100 samples, we create a sub-sample of 3550 TNG50 objects that mimics the TNG100 sample by selecting all TNG50 galaxies from the snapshots of the TNG100 sample (Fig. 1), whose

stellar mass is above the TNG100 mass threshold, $10^{9.5} M_{\odot}$. The distribution of scores for this sub-sample is shown in the lower panel of Fig. 6 as a black dotted line. While the distribution for this sub-sample is more similar to TNG100's, it is still skewed to higher scores in comparison. The fraction of JF drops from 6.9 per cent in the full TNG50 sample to 4.8 per cent in the TNG50 sub-sample, which is still larger than the TNG100 sample fraction of 3.1 per cent. This suggests that the enhanced resolution of the TNG50 simulation is indeed more conducive to identifying the salient features of JF galaxies.

Fig. 4 shows 20 random examples of galaxy images identified as JF galaxies, i.e. with a score of 0.8 and above.

3.2 Comparison to previous classification projects

To assess the public classifications of the CJF project, we compared them with classifications completed by a team of experts. This allows us to assess the extent to which the classifications from the general public align with our understanding of what comprises a JF. We compare results for a subset of objects that have also been independently classified by members of our research team. In particular, we make two comparisons: (1) against a subset of TNG50 objects classified in a pilot project completed *prior* to the public opening of the CJF Zooniverse website and (2) against the galaxies inspected and studied in Yun19.

3.2.1 Comparison with the TNG50 pilot project

Our in-house pilot study is functionally identical to the final CJF project. The sample includes 3466 objects from the TNG50 simulation, from the $z = 0, 0.2, 0.5, 1, 1.5$, and 2 snapshots (541, 585, 690, 673, 552, and 425 objects, respectively): This is a subset of the CJF Zooniverse project of Figs 1, 2, and Table 1. The classification team consists of six team members, five of which also classified for Yun19.

In Fig. 7, left panels, we show a comparison of the results and the raw JF scores for the CJF Zooniverse project and our pilot project (both normalized to values in the range 0–1). In Fig. 7(a), the numbers in each score bin show the percentage of the sample. Values along the secondary diagonal (bottom left to top right) show the percentage of objects for which there is complete agreement (within the shown bins). Summing along the diagonal we find complete agreement for about 84 per cent of the objects. To answer the binary choice: ‘Is the galaxy a JF, yes or no?’, we consider the collapsed matrix, with all scores below 0.8 considered to be not a JF: This is given in Fig. 7(c). We find that for the question of whether or not a galaxy is a JF, there is an agreement of 96.5 per cent of all objects. Most of the objects with inconsistent outcome are considered JF by the experts but not by the general public. The degree of agreement remains the same for JF thresholds of either 0.66 or 0.9.

The high degree of agreement shows that the classification by non-experts, with a high enough number of classifications per object, is a viable alternative to expert classification for the purpose of identifying JF galaxies in gas maps.

3.2.2 Comparison with the Yun19 Project

Of the sample used and studied in Yun19, 2142 galaxies are included in the CJF Zooniverse project: these are all from the TNG100 simulation. The Yun19 classifications use galaxy images that are similar but not identical to those in the subsequent TNG50 pilot

and CJF projects, and without a dedicated common platform for classification. In both cases, images are based on a combination of a colour map for gas column density and of stellar mass contours (see Section 2.3.1), with the same density limits. However, the smoothing procedure is not necessarily identical. Furthermore, a major difference between the images is the background subtraction in the Yun19 project. In that case, two side by side images of gas density for each galaxy, with one of the images subtracted by the mean gas density, enhances identification of features such as gas tails and bow shocks (see fig 1. of Yun19).

Fig. 7, right panels, summarizes the comparison for the commonly inspected images in the CJF Zooniverse and Yun19 projects. Summing along the diagonal of Fig. 7(b), we find complete agreement for about 41 per cent of the objects. Most of the discordant cases are for galaxies for which the experts give a higher JF score than the volunteers – consistent with the advantages inherent in the background-subtracted images. The collapsed matrix for the binary states (JF versus non-JF) shown in Fig. 7(d) gives an agreement for 82.3 per cent of inspected and common galaxies.

We speculate on the source of this disagreement, and why the outcome is so different than for the comparison with the TNG50 pilot project. As noted above, the Yun19 project used different images for the classification, including images with background subtraction. In addition, the Yun19 project was the first classification campaign for the team, and it is possible that with increased experience, the subjective criteria for identifying JF might have refined and improved. Finally, the TNG50 pilot project uses a setting almost identical to that of the CJF campaign, with respect to the images, and the classification platform. In contrast, for the Yun19 visual inspection, the classifiers all use the same images, but view them in disparate ways.

Despite the differences between the Yun19 and CJF classifications, the overall agreement is greater than 80 per cent. In both comparisons, we find that most objects for which there is disagreement are of the ‘false-negative’ variety, i.e. galaxies that experts see as JF but are not identified as such by the general public, suggesting that the JF population identified by the general public is pure but perhaps not complete. Overall, we find that we can trust volunteer classifications so long as a sufficiently large number per object are available.

3.3 Adjusted JF scores

Given the initial analysis of the visual inspections discussed above and the resultant galaxy scores, we propose to adopt a more nuanced interpretation of the CJF Zooniverse classifications. To do so, we assess the expertise of each inspector, and assign a weighting accordingly (e.g. Lintott et al. 2008).

As could be seen in Fig. 5, roughly 19 per cent of inspectors classify 10 images or fewer. As with any learned task, increased experience (usually) leads to higher proficiency, and this should be reflected in assessing the classifications. Conversely, there are several hundred inspectors who have classified more than a thousand images, with some having classified more than 10^4 images (see Section 2.3.2). These inspectors are responsible for more than half of the total classifications, and appraising the quality of their classifications is also important.

One issue which may impair our inspector weighting scheme is the issue of classifications by unidentified inspectors. There is no way to generate and evaluate the voting record of an inspector if they are identified differently in each session. We suspect that in many cases, a non logged-in user is someone who only classified a few images

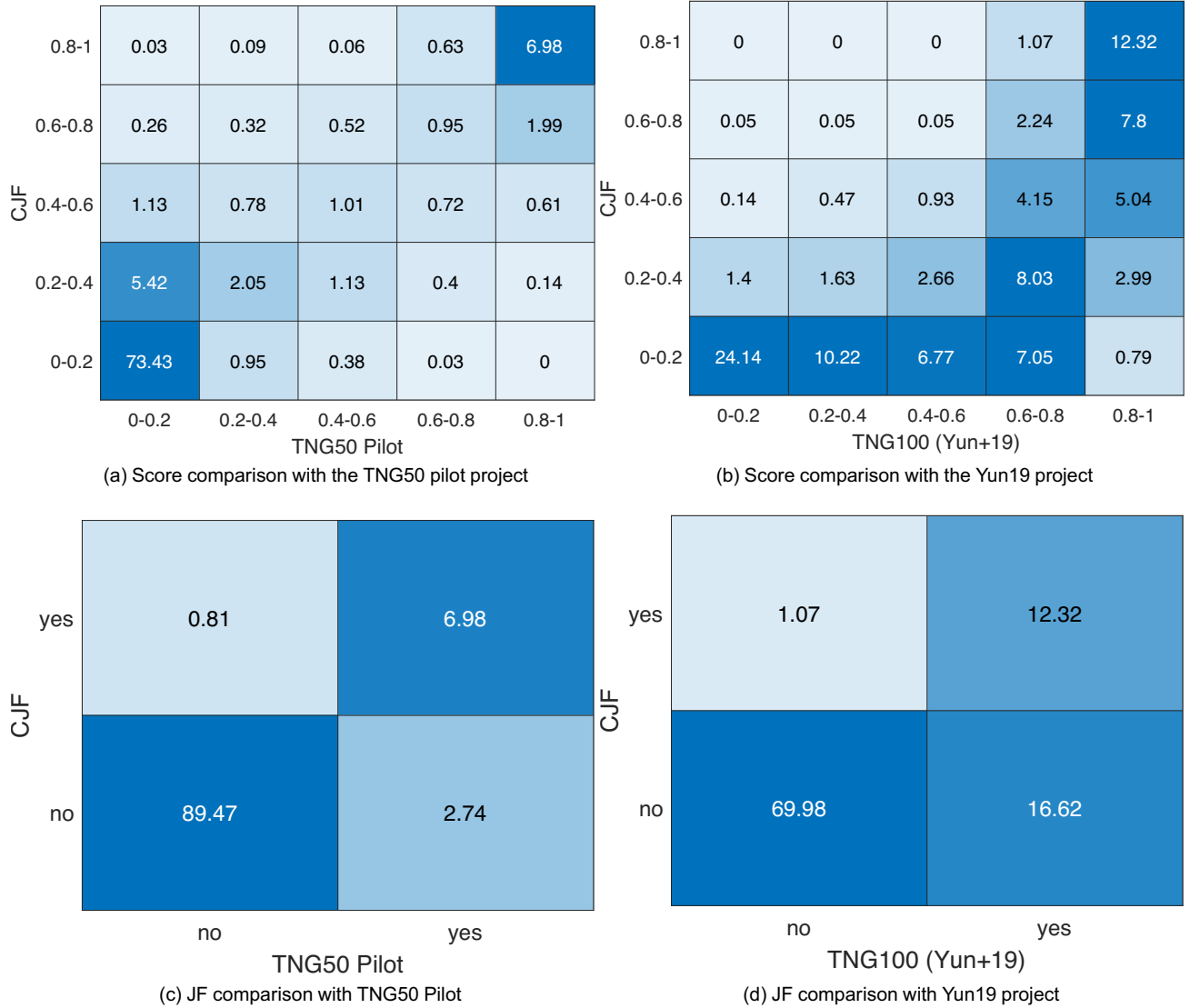


Figure 7. Comparison of the C/JF Zooniverse results to previous studies carried out by expert inspectors from the research team on a subset of objects. We compare to a pilot project on TNG50 galaxies (Figs 7a and c) and to the Yun19 TNG100 sample and classifications (Figs 7b and d). The top rows show the raw score comparison in five score bins, with the bottom rows showing the ‘collapsed’ comparison for the binary options JF/non-JF (corresponding to raw scores above/below the value of 0.8). The TNG50 pilot, performed on identical images and using the same C/JF Zooniverse platform, shows very high agreement. The agreement is not as strong for the comparison with the Yun19 visual inspections, most likely due to differences in the adopted images.

before growing dis-interested and moving on. For these cases, the inspector-weighting scheme works as intended.

Finally, we wish to take advantage of the classification of objects by experts, i.e. members of the team with experience in image classification in previous projects, both for JF identification or other galaxy-inspection tasks.

To this end, we assign weights to individual inspectors based on their experience and voting history. The revised score of an object is then set to be

$$\text{score} = \frac{\sum w_i v_i}{\sum w_i}, \quad (1)$$

where the summation is over all inspectors who classify the object, w_i is the inspector weight and v_i is the vote (0,1) given by the inspector. The inspector weight is set by the following scheme:

(i) Inexperienced inspectors that have classified fewer than 10 objects are all given a weight of 0.5.

(ii) Expert inspectors who are members of the research team are all given a uniform weight of 5. The classifications from the TNG50 Pilot and the Yun19 projects are incorporated into the final score as additional expert classifications.⁵

(iii) Performance on high-score objects. If an inspector votes against the consensus from their peers on high-score objects, their weight is reduced. This favours high-accuracy inspectors.

(iv) Removal of repeat offenders. Inspectors who consistently misidentify JF galaxies, voting ‘no’ on objects that received predominately ‘yes’ votes (down-voters) and vice versa (up-voters), are given a weight of 0, i.e. removed completely. The number of inspectors removed in this manner is 338 (125 down-voters and 213 up-voters).

⁵Using the adjusted scores, the agreement between the C/JF project and the TNG50 pilot project is now 98.9 per cent, and the agreement with the Yun19 sample is 93.7 per cent (Sections 3.2.1 and 3.2.2).

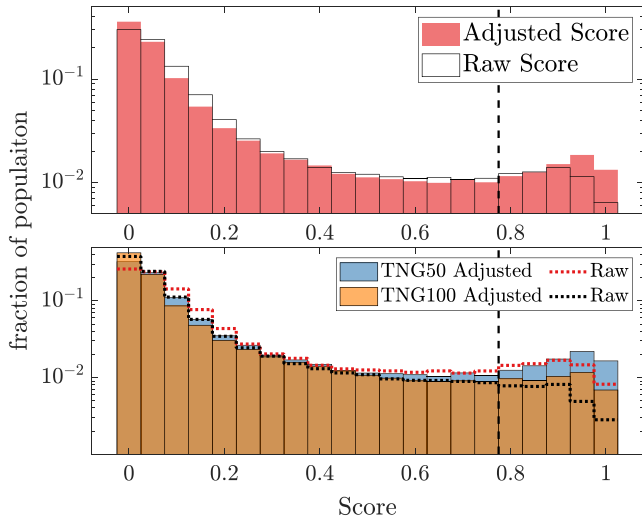


Figure 8. Comparison between raw and adjusted JF scores. The distributions of the raw and adjusted JF scores for the entire inspected sample (TNG100 + TNG50 galaxies) are shown in the top panel, with the numbers in each bin denoting the percentage of the population. The bottom panel shows the adjusted score distribution for the TNG50 and TNG100 samples separately (blue and orange bars, respectively), with the raw score distribution for the two samples shown by the red and black dotted lines. The JF threshold of 0.8 is marked by the dashed vertical line. Due to the score adjustment, an additional ~ 1 per cent of the population is identified as a JF bringing the total number of JF from 4553 to 5307.

The inspector weighting algorithm is detailed in Appendix A. With this approach, some objects receive scores based on less or more than 20 votes. However, the final scores of all objects are determined by the votes of at least 13 inspectors, and in 95 per cent of cases the score is determined by 18 or more inspectors.⁶

In Fig. 8, we show the histogram of the adjusted JF scores compared to the raw scores in the top panel, with the scores of the TNG50 and TNG100 samples shown separately in the bottom panel (a revised version of Fig. 6). As can be seen, at the low-score end the adjusted scores are lower than the original, i.e. raw scores, and conversely, at the high-score end the adjusted scores are higher. Fig. 9 shows a detailed comparison between the raw and adjusted scores for the entire inspected sample. Most objects in the sample are found close to the diagonal showing that their scores do not change much due to the weighting scheme. When scores do change due to the weighting scheme, it is usually towards higher scores, as intended.

Due to the score adjustment, an additional ~ 1 per cent of the entire population is identified as a JF galaxy, as defined in Section 4.1. This results in an increase of 37.5 per cent in the number of TNG100 galaxies identified as JF, and an increase of 11.8 per cent for TNG50 galaxies (an increase of 16.6 per cent overall). Namely, adopting the adjusted scores instead of the raw ones returns a total population of IllustrisTNG JF comprised of 5307 objects instead of 4553.

The reason for the large increase of nearly 40 per cent of JF in the TNG100 sample due to the adjusted score, compared to only 12 per cent in the TNG50 sample, is most likely due to the

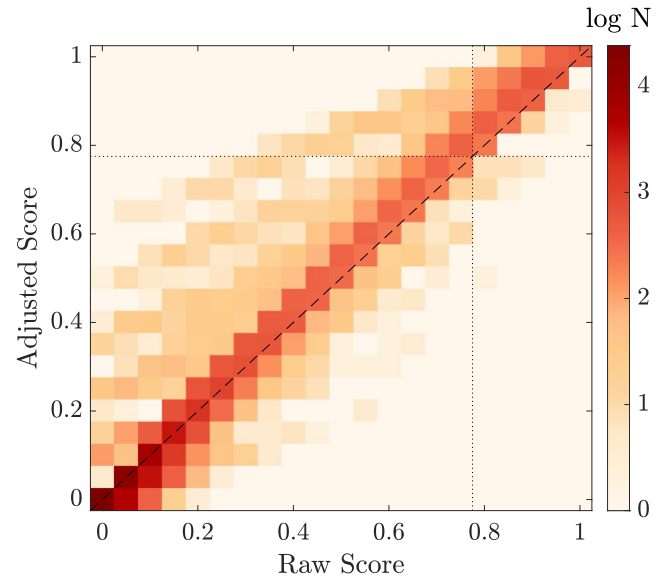


Figure 9. A one-to-one comparison between the adjusted and raw scores for the entire inspected galaxy sample (TNG100 + TNG50 galaxies). The dotted horizontal and vertical lines mark the JF threshold.

contribution of expert votes to the weighting scheme which were included from the Yun19 sample: of all the TNG100 objects whose raw score was below the 0.8 JF threshold and adjusted scores was above it, 74 per cent were objects also studied in the Yun19 project. In comparison, the fraction of such objects in the TNG50 sample is only 11 per cent. As detailed in Section 3.2.2, there was a marked discrepancy between expert and non-expert classification of the Yun19 project sample, most likely due to the difference in the images used for classification.

4 GUIDELINES TO DECIDE JF STATUS

One of the ways in which we envision the use of this inspected sample of galaxies, each with their JF score (adjusted or not), is by establishing an appropriate threshold above which galaxies are considered JF.

In Fig. 10, we present a sample of 20 images of galaxies that are organized into five equally spaced score ranges, as assigned by the classification scheme and adjusted as described above. In the top row, we show the high-score objects (0.8–1.0), whereas lower rows show progressively lower score ranges, with the bottom-most showing objects of score 0–0.2. Additional, randomly selected examples of images in these score ranges can be seen in Appendix B.

Based on these images, we find that, as expected, objects in the highest score bin all appear to be JF galaxies (see also Fig. 4), but some objects in the next score bin (0.6–0.8) could also be considered as JF, as described by the guidelines presented in the classification project (Section 2.3.2). Objects in the low ranges indeed do not resemble JF galaxies. Our fiducial threshold value throughout the paper is 0.8, as we expand upon below.

4.1 JF threshold in this study

In this work, we adopt and hence suggest a fiducial (adjusted, inspector-weighted) threshold score of 0.8 or better to consider a galaxy as a JF. This choice is driven by our requirement for a pure sample population, i.e. one that we could be sure to contain, to a good

⁶The details of the weighting scheme are chosen to enhance the identification of JF galaxies to enable an analysis of the demographics of the JF galaxy population. Future research questions may require a different approach, and a different weighting scheme altogether, and we encourage future users of these data sets to consider whether and how they wish to formulate a score that best fits their research question.

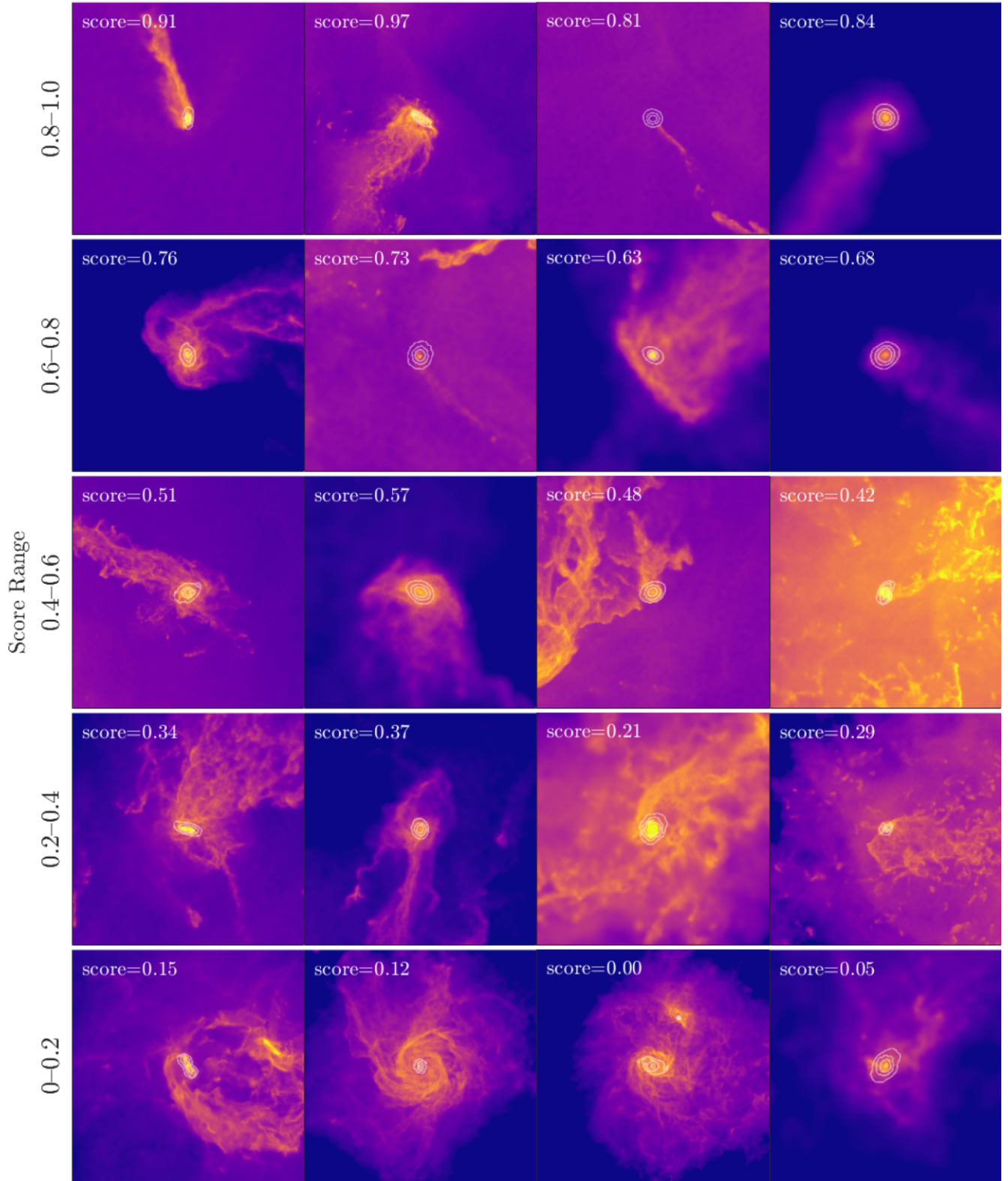


Figure 10. Examples of IllustrisTNG galaxy images of varying JF scores according to the adjusted results from the CJF Zooniverse project. The top row shows objects of the highest scores, 0.8–1.0, with the lower rows showing objects of progressively lower scores. The score each image received appears in the top-left of each image. The image specifications are the same as in Fig. 4. More examples can be found in Appendix B.

degree of confidence, very few non-JF galaxies, but that would also be large enough (several thousand objects) to constitute a representative sample of JF across different properties and environments. This requires a relatively high threshold, that may exclude some JF galaxies from our samples (e.g. second row of Fig. 10). Visual inspection of the fiducial IllustrisTNG JF sample convinces that the final sample is indeed comprised almost entirely of JF galaxies.

Our choice of JF score threshold is based on the experience gained from the previous pilot projects (Section 3.2) and on the considerations above. Furthermore, it is made to suit the needs of our research questions: we encourage future researchers to formulate their own threshold value according to the requirements of the scientific goal in question. Below, we give additional motivations and guidelines on how to determine an optimal threshold.

4.2 Determining the JF score threshold

We now suggest a statistical framework to assign a confidence level for the choice of a threshold score, and to compare against different threshold choices. We make the following two assumptions in order to define a relationship between an image of a galaxy and its score in the project:

- (i) For each galaxy *image* there exists a probability j_p that an inspector will classify the image as a JF galaxy.
- (ii) The probability j_p is defined as the fraction of N inspectors who classify an image as a JF galaxy as $N \rightarrow \infty$.

Based on this definition, for a total number of N inspectors ($N = 20$ in our case), the conditional probability that an image with a given j_p will receive a score k is given by the binomial distribution

$$P(k, N | j_p) = \binom{N}{k} j_p^k (1 - j_p)^{N-k}. \quad (2)$$

Using Bayes Theorem, we can state the more interesting probability that an image which received a score k has an inherent classification probability of j_p

$$P(j_p | k, N) = \frac{P(k, N | j_p) P(j_p)}{P(k, N)}. \quad (3)$$

Since we have no prior knowledge, we choose a flat prior for j_p ($P(j_p) = \text{const.}$) and the evidence $P(k, N) = \text{const.}$, thus we find

$$P(j_p | k, N) = \mathcal{A} \binom{N}{k} j_p^k (1 - j_p)^{N-k} \quad (4)$$

where \mathcal{A} is a normalization which ensures $\int_0^1 P(j_p | k, N) dj_p = 1$. This defines the probability density function (PDF) for the values of j_p as captured by the score from $N = 20$ inspectors.⁷

From this PDF, we define the cumulative density function (CDF) which gives the probability that j_p has a given value or less for a galaxy image, given a score of k out of N

$$P(\leq j_p | k, N) = \int_0^{j_p} P(j | k, N) dj. \quad (5)$$

Thus, the value $1 - P(\leq j_p | k, N)$ gives the probability that, given a score k , an image is characterized by a value of j_p or above, which is what we look for in a threshold defining a JF galaxy. Furthermore, this value constitutes the *confidence level* associated with a value of j_p or above of the galaxy image.

⁷The PDF of j_p can be used to explore how a larger or smaller number of inspectors affects our ability to reconstruct j_p .

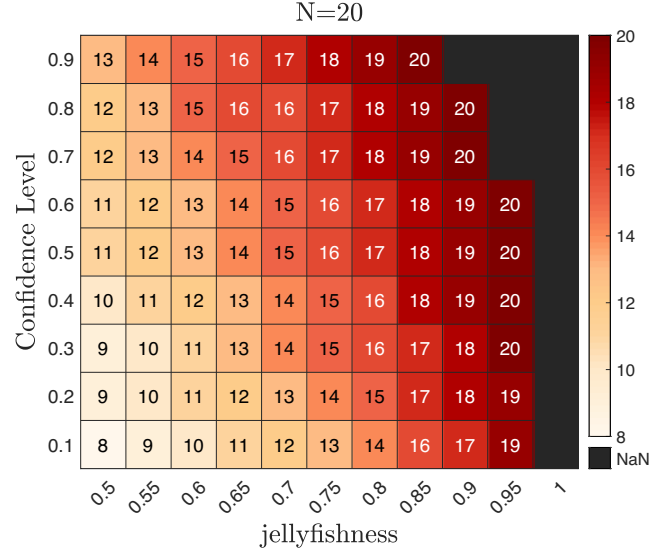


Figure 11. The relationship between raw JF scores from the CJF project (numbers in the cells), the classification probability of being a JF (j_p or ‘jellyfishness’) and the confidence level for the probability values. The confidence level refers to how well a classification probability of j_p or above can be ensured given a score from the $N = 20$ inspectors. The black region corresponds to a region which is not obtainable with only 20 inspectors.

In summary, for an image that receives a score k , we can not only attach one or more j_p values, but also assess the confidence level of said values. We demonstrate this in Fig. 11, where for each value of j_p (on the x -axis) and desired confidence level (on the y -axis), the number in the intersecting square shows the minimal score in the CJF project that ensures these values. For example, to guarantee a value of $j_p \geq 0.8$ at a confidence level of 80 per cent one must choose galaxies with a score of 18 or above, but for a confidence level of 40 per cent, a score of 16 will suffice. In addition, Fig. 11 can also be used to find the different j_p values associated with a given score, as well as the confidence level of these values.

We note that the probability j_p is associated with an *image* of a galaxy within the context of a given project, i.e. based on the image generation technique, instructions for inspectors, and so on. In addition, this model assumes equally weighted inspectors, but can be extended to address an inspector-weighting approach, as in Section 3.3 (e.g. Benneyan & Borgman 2004). However, this would only increase the confidence level relating a score and j_p value.

In the case of our fiducial approach, for 20 inspectors, the advocated threshold of 0.8 (16 ‘yes’ votes) ensures that all our images are within the top quartile of ‘jellyfishness’ ($j_p \geq 0.75$) with a high level of confidence (larger than 60 per cent, thanks to the inspector-weighting).

5 ILLUSTRITNG JF: BASIC DEMOGRAPHICS

Based on the score distribution of the galaxies in the inspected sample as shown in Fig. 8, we see that JF galaxies comprise a small percentage of the population of inspected satellites (and an even smaller one of the total satellite population) simulated within IllustrisTNG. For most galaxies, there is little doubt that they are *not* JF galaxies: ~ 44 per cent (42 per cent in TNG50 and 51 per cent in TNG100) have a score of 0.05 or less, and ~ 93 per cent have a score of 0.25 or less (92 per cent in TNG50 and 96 per cent in TNG100).

Table 2. Number of visually identified JF galaxies, and the resultant JF fractions, in the TNG50 and TNG100 simulations, in three redshift bins and according to the adjusted scores from our CJF Zooniverse project.

Redshift	TNG50		TNG100	
	JF galaxies	JF fraction	JF galaxies	JF fraction
$0 \leq z \leq 0.5$	3971	8.2 per cent	1016	5.7 per cent
$0.5 < z \leq 1$	138	4.7 per cent	23	2.2 per cent
$1 < z \leq 2$	35	1.6 per cent	23	0.7 per cent
Total	4144	7.7 per cent	1163	4.3 per cent

Notes. The JF fractions (shown in percentage) are obtained by dividing the JF number by the total number of *inspected* satellites for a given redshift range (see Table 1).

The total number of JF galaxies identified in the TNG50 and TNG100 simulations via our CJF Zooniverse project is given in Table 2, where we also separate into redshift bins: 5307 JF in total and available for scientific studies. The total JF fraction out of the inspected satellite sample (score of 0.8 or above) is 6.6 per cent (7.7 per cent in TNG50 and 4.3 per cent in TNG100). Most of the JF galaxies are found at low redshifts, and the JF fractions are also highest at those times. However, interestingly and as we expand upon in the next sections, even at redshifts between $z = 1$ and 2 JF galaxies comprise ~ 1 per cent of the inspected satellite population.⁸

In the following sections, we explore the demographics of JF galaxies by examining the number and frequency of IllustrisTNG JF galaxies in relation to satellite stellar mass, host mass, and redshift.

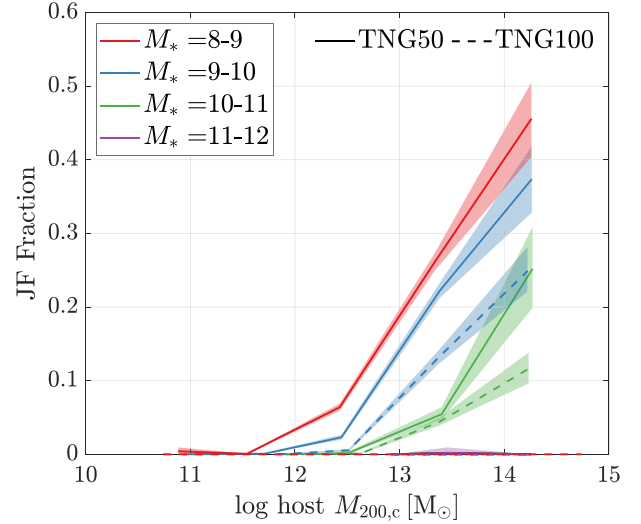
5.1 Demographics of the JF population and their hosts

Since there is a large difference in the TNG50 and TNG100 selected samples in terms of satellite stellar mass, host mass, and redshift ranges (see Figs 1, 2, and Table 1), we present the fraction of JF separately for the two samples. To assess the spread in values of the JF fractions, we use the bootstrap method within a sub-sample of galaxies (defined by a combination of satellite stellar mass, host mass, redshift, etc.) and show the resulting 95 per cent confidence level. We note that these fractions relate the number of JF galaxies compared to the entire inspected satellite sample, and not the entire satellite population found in the simulations. However, the change in fractions is at most a factor of 2 at low redshifts, and less at higher redshifts, based on the total number of satellites in the relevant mass range in TNG50 and TNG100 (see Section 2.2).

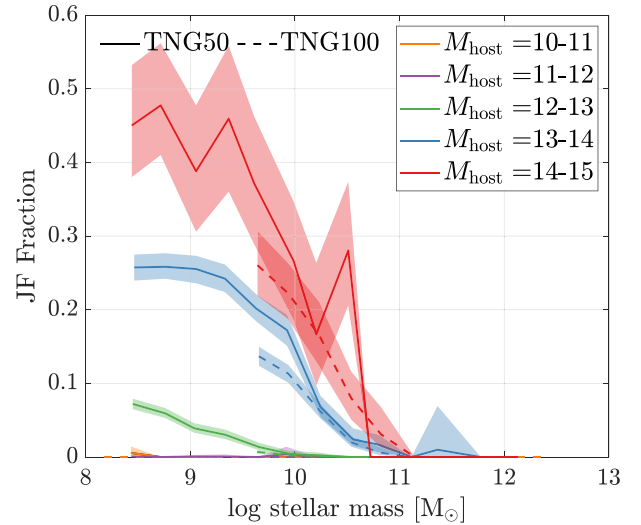
5.1.1 Demographics of JF galaxies

In Fig. 12, we examine the JF fractions in bins of satellite stellar mass and host mass. We bin the galaxies in four equally sized bins in log stellar mass: $10^8 - 10^9$, $10^9 - 10^{10}$, $10^{10} - 10^{11}$, and $10^{11} - 10^{12} M_\odot$, and five equally sized bins in log host mass ($M_{200,c}$): $10^{10} - 10^{11}$, $10^{11} - 10^{12}$, $10^{12} - 10^{13}$, $10^{13} - 10^{14}$, and $10^{14} - 10^{15} M_\odot$. We refer to the latter bin as cluster-sized hosts and notice that, strictly speaking, this is populated not all the way up to $10^{15} M_\odot$ but up to $10^{14.6} M_\odot$, the most massive cluster in TNG100 at $z = 0$.

⁸These JF fractions are lower than those found in Yun19. While here JF fractions are of order a few per cent, in Yun19 the total JF fraction over the entire inspected sample was 13 per cent. However, there is a large difference in the two samples, in terms of satellite stellar mass, host mass, and redshift. When comparing the JF fractions under the same selection restrictions for the CJF inspected sample, we find a JF fraction of ~ 18 per cent.



(a) JF fraction vs. host mass in satellite stellar mass bins



(b) JF fraction vs. satellite stellar mass in host mass bins

Figure 12. The JF fraction according to the IllustrisTNG simulations as a function of satellite stellar mass and host mass bins. The JF fraction is the percentage of JF to the inspected satellite number in each stellar/host mass bin (not all satellites have been visually inspected – see sample selection criteria in Section 2.2). In panel (a), we show the JF fraction versus the host mass ($M_{200,c}$) in four different bins in log stellar mass. The JF fraction versus the satellite stellar mass is shown in (b), divided into five host mass bins. Solid lines show the fractions in the TNG50 sample and dashed lines correspond to the TNG100 sample. The shaded regions denote the 95 per cent confidence levels intervals based on bootstrapping within each host mass/stellar mass bin. Objects from all redshifts are included here. JF fractions are higher for lower mass galaxies and in higher mass hosts.

Due to the resolution-dependent mass threshold, the lowest satellite stellar mass bin is not populated in the TNG100 sample, and as a result, the two lowest host mass bins are empty in this sample as well (Fig. 2). The highest host mass bin in TNG50 is populated by only 29 objects, most of which are likely different instances of only 1 or 2 clusters in different snapshots.

In Fig. 12(a), we show the fraction of JF galaxies out of the inspected satellite population, over all the hosts in a given mass bin. This includes *all* hosts, and not only those that actually contain JF,

which are a small minority of the host population. We focus instead on the JF populations of individual hosts in Section 5.1.2.

We see that the JF fractions grow with increasing host mass. In hosts of masses up to several times $10^{12} M_{\odot}$ a few per cent of satellites are JF, and all of these are necessarily low-mass satellites. However, this frequency grows to 10–20 per cent in group-mass hosts ($\sim 10^{13} M_{\odot}$) and up to 40 per cent for low-mass satellites in cluster-mass ($\sim 10^{14} M_{\odot}$) systems in TNG50. The TNG100 values are lower, but still reach values of 10–20 per cent in clusters.

These JF fractions are qualitatively consistent with recent observations that probe the group-mass scale (Roberts et al. 2021a). Furthermore, this finding explains why the JF fractions are higher at lower redshifts (Table 2): The number of high-mass hosts increases towards lower redshift, supplying an environment more conducive to the formation of JF galaxies. Additionally, the longer the time galaxies spend in high-dense environments, the higher are the chances for them to undergo RPS.

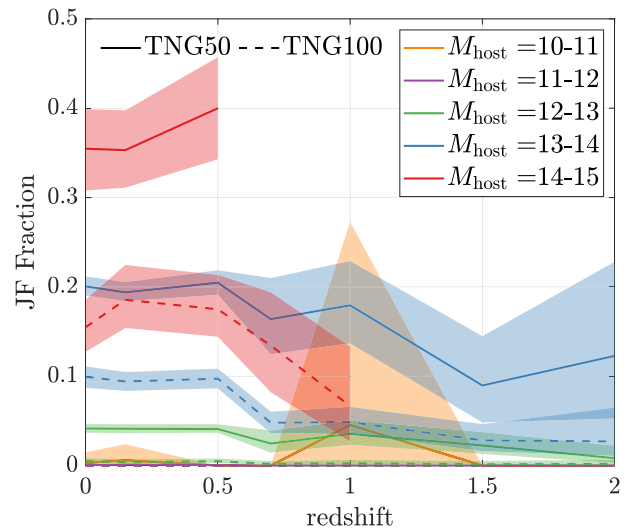
On the other hand, while rare, JF galaxies can be found even in hosts of mass $\sim 10^{12} M_{\odot}$, namely around galaxies of mass similar to our own Milky Way and Andromeda. Additional considerations on the presence of JF galaxies around TNG50 Milky Way and M31-like galaxies can be found in Engler et al. (2023).

In Fig. 12(b), we show JF fraction versus satellite stellar mass, separated in halo mass bins.⁹ The TNG100 sample, which includes roughly three times as many objects, exhibits much smoother values. We see that the JF fraction drops with increasing satellite stellar mass. The more massive the galaxies, the harder it is to remove gas by RPS. For low-mass galaxies ($< 10^9 M_{\odot}$), JF comprise more than 40 per cent of the inspected satellite population in cluster-sized systems, and nearly 30 per cent in groups. These results explain the higher JF fractions found in the TNG50 sample compared to the TNG100 galaxies: due to the higher resolution, we can study and hence have many more low-mass galaxies in the TNG50 sample (see selection criteria in Section 2.2).

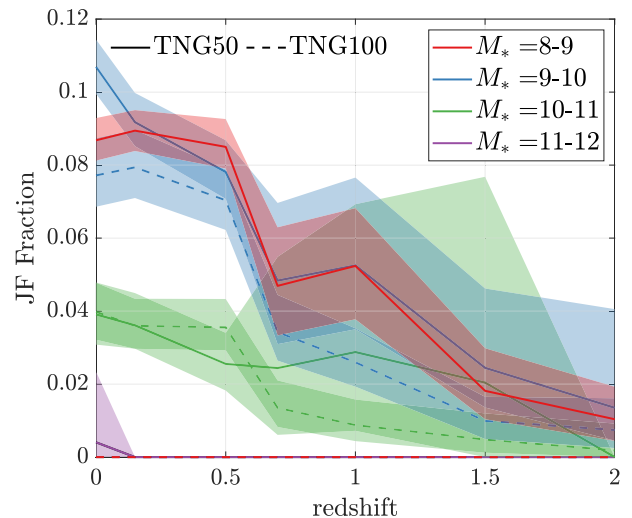
In Fig. 13, we explore how the JF fractions change with redshift, at fixed host mass and satellite stellar mass bins. In Fig. 13(a), we see that the first high-mass clusters only appear after $z = 1$ in TNG100 and $z = 0.5$ in TNG50. The large difference in the JF fraction between the two samples is due to the different mass-cut employed in the samples: The TNG100 sample does not contain galaxies below $10^{9.5} M_{\odot}$, where the JF is much higher (Fig. 12b).

We see that JF galaxies can be found as early as $z = 2$, mostly in groups and protoclusters of mass 10^{13} – $10^{14} M_{\odot}$, where they account for ~ 10 per cent of all the satellites in these hosts. Even in hosts with masses of $10^{12} M_{\odot}$ we find JF galaxies at these high redshifts. The JF fraction increases with decreasing redshift, by roughly a factor of 2 between $z = 2$ and 0. The JF fractions in low-mass hosts of $\sim 10^{10} M_{\odot}$ are tenuous: we find only a single such JF galaxy in the TNG50 sample. In fact, there are only four objects identified as JF galaxies found in hosts of mass $\sim 10^{10} M_{\odot}$, which we study in detail in Section 5.2.3. The bootstrap uncertainty estimates produce large shaded regions for bins containing few galaxies.

The frequency of JF in satellite stellar mass bins also rises with decreasing redshift, as shown in Fig. 13(b). The evolution of the JF fraction is similar in the two samples. At the high-redshift end, the JF fractions are roughly 1 per cent for all mass bins, except for the most massive galaxies which are rare at these epochs (only 24 galaxies of



(a) JF fraction vs. redshift by host mass



(b) JF fraction vs. redshift by satellite stellar mass

Figure 13. The redshift evolution of the frequency of IllustrisTNG JF galaxies in host mass bins (a) and satellite stellar mass bins (b). Annotations are as in Fig. 12.

mass $10^{11} M_{\odot}$ and above for $z > 1$ in the entire inspected sample, none of them JF). The JF fractions rise with decreasing redshift in a similar manner for the lower mass satellites ($\leq 10^{10} M_{\odot}$), reaching values of 8–11 per cent.

In the higher mass satellite mass bin, 10^{10} – $10^{11} M_{\odot}$, the increase is milder and only reaches 4 per cent or so. In fact, the most massive satellites (with stellar masses of 10^{11} – $10^{12} M_{\odot}$) represent a subset of particular interest and complexity: they are JF galaxies only at very low redshifts. In fact, we find only one JF galaxy of mass exceeding $10^{11} M_{\odot}$ (TNG50 at $z = 0.14$), among 1205 inspected ones. A number of physical processes are at play in the case of these massive galaxies. On the one hand, higher-mass galaxies exert a stronger gravitational pull and, as such, it stands to reason that there should be fewer JF galaxies at these mass ranges. Additionally, as shown, for example, by Terrazas et al. (2020) and Zinger et al. (2020), at a stellar mass of $\gtrsim 10^{10.5} M_{\odot}$, the AGN kinetic feedback in IllustrisTNG becomes important and can evacuate most of the gas from the inner regions, or even haloes, of galaxies. This may also contribute to the relatively

⁹The fluctuations in clusters in the TNG50 sample are due to small number statistics. There are only two $\sim 10^{14} M_{\odot}$ hosts in TNG50 at low redshifts (Joshi et al. 2020).

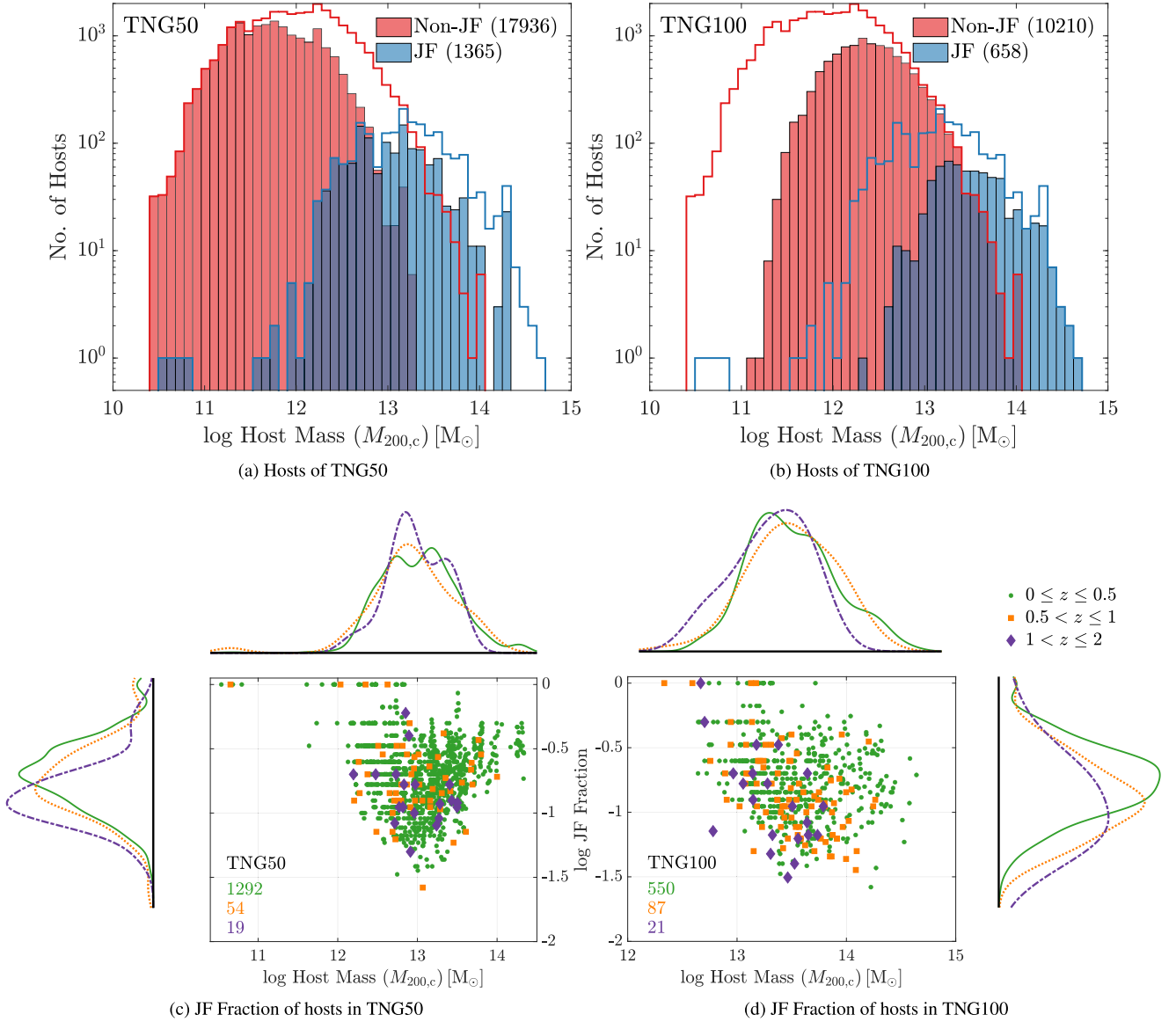


Figure 14. Demographics of IllustrisTNG hosts that contain JF galaxies. Histograms of the number of hosts of a given $M_{200,c}$ separated into those that contain JF (in red) and those that do not (in blue) are shown for TNG50 and TNG100 in panels (a) and (b), respectively. The empty histograms show the distribution of the combined inspected sample. For all hosts that contain at least one JF satellite, we show the JF fraction versus the host mass $M_{200,c}$ in the TNG50 and TNG100 samples in the lower panels, panel (c) and (d), respectively. Hosts are colour-coded into three redshift bins: $z \leq 0.5$ (green), $0.5 < z \leq 1$ (orange), and $z > 1$ (purple). The total number of hosts in each panel appears in the bottom left corner. Normalized histograms of the projected distribution along the axes are also shown.

lower number of JF galaxies in this mass range. On the other hand, we have shown that, according to IllustrisTNG, the AGN feedback in massive satellite galaxies is generally hampered by their high-density environments compared to that in similar-mass field galaxies (Joshi et al. 2020). Furthermore, it is not clear whether AGN-driven outflows hinder the emergence of a JF phase by completely removing gas or whether they could actually promote RPS and hence the JF outlook of galaxies by making the gas less gravitationally bound. We postpone to future work more detailed investigations on this.

5.1.2 Demographics of hosts that contain JF

In this section, we focus on the demographics of the host haloes that contain JF galaxies. In Figs 14(a) and (b), we show the distribution

of $M_{200,c}$ of all hosts that host the galaxies in the inspected sample, separated by simulation. In each simulation sample, we show the distribution of the haloes that host JF galaxies in red, and the hosts in which no JF galaxies were found in blue. The empty histogram in each panels show the distribution of the combined TNG50 and TNG100 samples, and is thus identical in both panels. It should be noted that, since the inspected sample spans multiple snapshots, there are different instances of the same objects included in each sample (see Section 6.2 for more details).

In general, higher mass hosts are more likely to have JF galaxies, and most haloes of $10^{13.5} M_\odot$ and above contain at least one JF satellite. Above a certain mass, *all* the hosts in the two samples contain JF galaxies: In TNG50, all hosts above $10^{13.3} M_\odot$ have a JF satellite, while in TNG100 all but three hosts above $10^{14} M_\odot$ contain

JF galaxies. This difference is due to the difference in the underlying numerical resolution and the correspondingly chosen satellite mass threshold in the two simulations.

Conversely, lower mass hosts are less likely to host JF galaxies, though this too is a simulation-dependent statement. Because of the minimum stellar mass threshold, the lowest mass haloes to host a JF galaxy in TNG100 are of mass $\sim 10^{12.5} M_\odot$. In TNG50, we find four low-mass hosts, $\sim 10^{10.5} M_\odot$, with JF galaxies (one each), with the next massive host with JF satellites is found at $\sim 10^{11.5} M_\odot$. As we discuss in Section 5.2.3, of these four JF galaxies, three may be linked to the wrong host and one may be mis-classified. As such, a more conservative estimate for the halo mass threshold for hosting a JF in the TNG50 sample is $\sim 10^{11.5} M_\odot$.

A more detailed view of the hosts of JF galaxies is shown in Figs 14(c) and (d) where we plot the JF fractions of individual hosts (as long as that fraction is non-zero), versus the host mass ($M_{200,c}$). Here, the JF fraction is the number of JF galaxies associated with the host divided by the total number of inspected satellites of the host. We separate TNG50 (left panel) and TNG100 (right panel), and split each into three redshift bins (three different symbols and colours).

Most hosts that contain JF galaxies are found at low redshifts: $z \leq 0.5$. However, there are tens of hosts containing JF galaxies at earlier cosmic epochs, up to $z = 2$. We note that some of these objects may be different instances of the same haloes in different snapshots; however, since each high redshift bin includes only two snapshots (see Fig. 1), the actual number of haloes containing JF galaxies can be smaller by a factor of 2 at most. Of the 65 hosts with a JF fraction of unity, 63 have one only inspected sample satellite, with the rest having only two or three satellites. The horizontal lines at certain JF fraction values are likely due to the quantized nature of the JF scores, combined with a small number of inspected galaxies in a given host, a situation more likely for low-mass hosts. Overall, the JF fraction distributions for the three redshifts are similar, between 10 and 30 per cent.

Now, the demographics considerations above do not allow us to answer the following question: based on the outcome of the IllustrisTNG simulations and of the CJF Zooniverse visual inspections, how many JF galaxies shall we expect to find in any given observed group or cluster of galaxies? In Fig. 15, we therefore present the statistics for the actual number of JF galaxies residing within $2R_{200,c}$ of individual hosts, across the usual five host mass bins: $< 10^{10} M_\odot$, $10^{11} - 10^{12} M_\odot$, $10^{12} - 10^{13} M_\odot$, $10^{13} - 10^{14} M_\odot$, and $> 10^{14} M_\odot$. We restrict the satellites to $2R_{200,c}$ to demonstrate the numbers of JF galaxies that may be reasonably expected to be found in observations of such hosts. Since the two simulations have different satellite stellar mass and host mass populations (see Fig. 2), we again present the JF statistics separately for TNG50 (red) and TNG100 (blue). The median number of JF satellites per host is indicated by a horizontal line, and the 25–75 per cent interquartile is shown by the box. Low-mass hosts ($10^{13} M_\odot$) with JF satellites are the rare exception, rather than the rule, but at group- and galaxy cluster-scales one can expect to find several JF galaxies. For example, we should expect to find 3 (0) JF galaxies with stellar mass above $10^{8.3} M_\odot$ ($10^{9.5} M_\odot$) in the typical Fornax-like group. However, depending on the state and assembly of the host, there could be systems that host up to 10–15 JF galaxies. The numbers are higher for the TNG50 inspected sample due to the lower satellite masses, which implies that probing even lower masses will yield higher numbers of JF satellites.

We note that these JF numbers per host are higher than the values found in observations of the LoTSS survey (Roberts et al. 2021b), but not inconsistent with them, given our much lower satellite

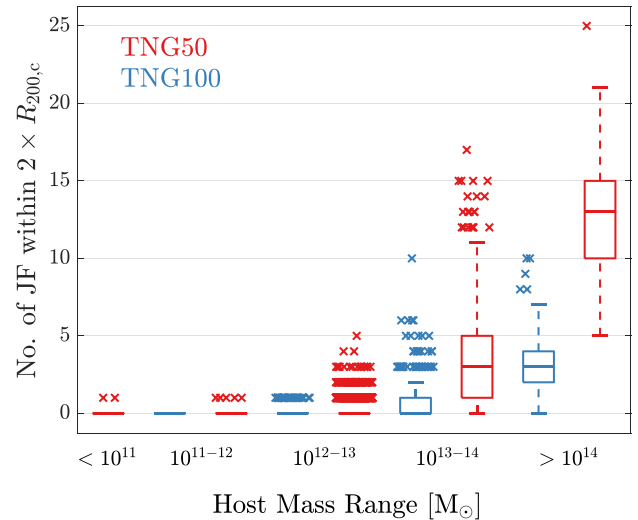


Figure 15. Number of JF galaxies to be expected in individual hosts within $2R_{200,c}$ and across host mass bins, according to the TNG50 and TNG100 simulations (red and blue, respectively). The 25–75 per cent interquartile range is shown by the box, with the median marked by the horizontal line. Outliers, defined as being found 1.5 times the interquartile range above/below the 75/25-quantile, are marked with crosses. The dashed lines extend to the farthest values which is not an outlier. There are no $< 10^{11} M_\odot$ hosts in the TNG100 inspected sample.

stellar mass threshold, $10^{8.3} M_\odot$ in our sample versus $10^{9.7} M_\odot$ in LoTSS.

5.2 Stellar mass, host mass, and radial distance of JF galaxies

Fig. 16 summarizes in one visualization the richness of phenomenology described so far: there we show the distribution of the TNG100 and TNG50 JF populations in terms of their stellar mass, host mass and radial distance, in units of $R_{200,c}$. The distribution of the non-JF galaxies is also shown for comparison, and populations from each simulation are separated (left versus right).

The distributions are shown in the satellite stellar mass – host mass plane (Figs 16a and b), the 3D radial position – satellite stellar mass plane (Figs 16c and d) and the 3D radial position – host mass plane (Figs 16e and f). The JF population is traced by the blue contours, which enclose the 25, 50 and 75 and 98 percentiles: objects beyond the outermost contour are indicated by blue triangles. The non-JF population is likewise shown by the red contours. The number of JF and non-JF galaxies are indicated by the blue and red numbers in the upper right corner. Histograms of the normalized distribution along each axis are also given.

As already quantified in previous Sections, Fig. 16 confirms that most of the JF galaxies reside in high-mass hosts, with roughly half the population inhabiting group- and cluster-sized hosts of mass $10^{13} M_\odot$ and above: The peak of the distribution is found at $\sim 10^{13.5} M_\odot$. Two opposite trends lead to this configuration: conditions for forming JF galaxies become more favourable with increasing mass (Fig. 12) while the abundance of hosts decreases with mass (Fig. 2b). This is in stark contrast to the non-JF population where most satellites are found in hosts less massive than $10^{13} M_\odot$, since there are many more of these objects in the inspected sample (Figs 2b and d).

In terms of satellite stellar mass, the distributions of the two populations are very similar. However, there is a relative deficit in

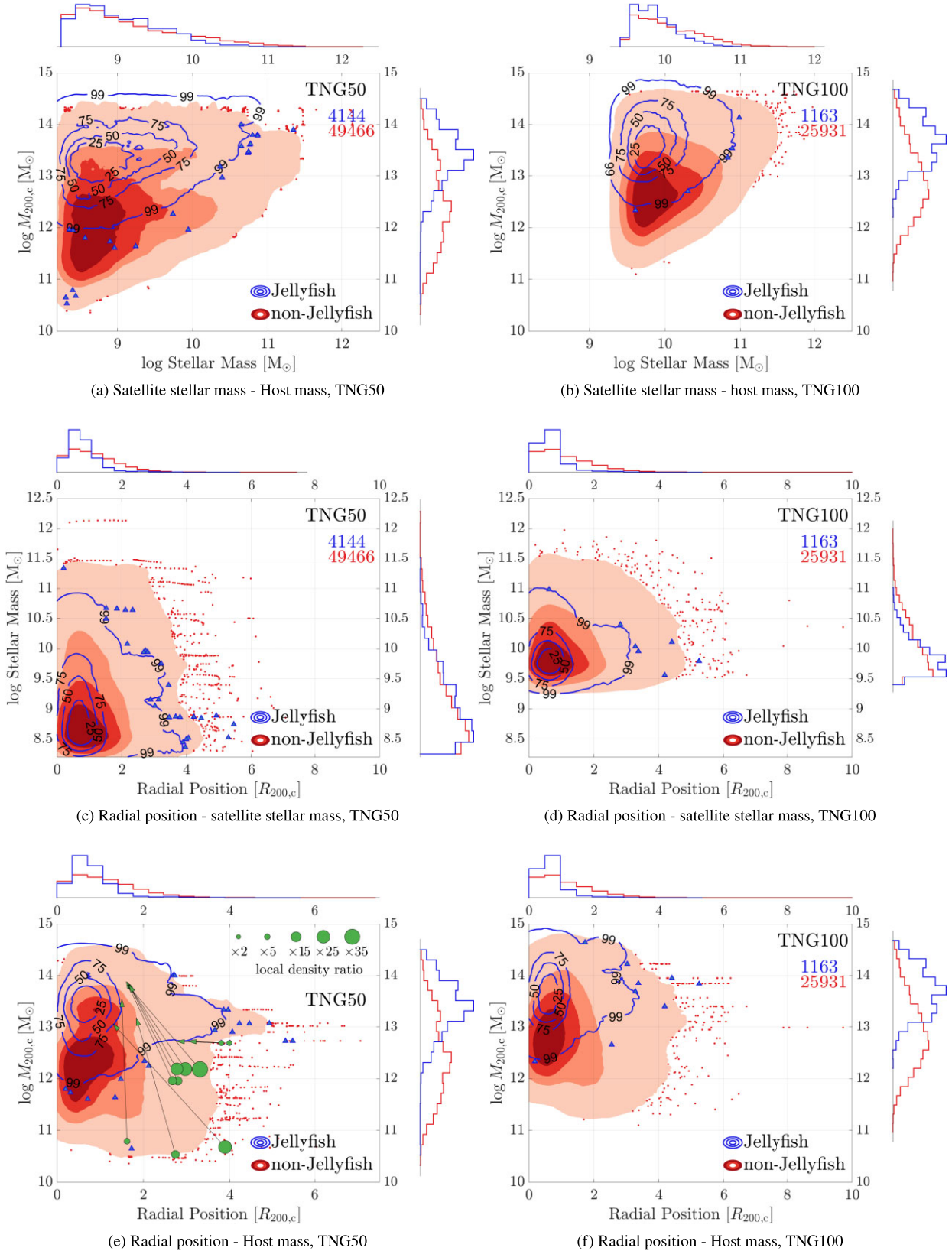


Figure 16. Distribution of inspected satellites and JF galaxies on the satellite stellar mass – host mass – radial position planes for TNG50 (left panel) and TNG100 galaxies (right panel). The radial position is the 3D distance to the centre of the host in units of $R_{200,c}$. JF galaxies are shown by blue contours enclosing 25, 50, 75, and 98 per cent of the JF population. JF galaxies found beyond the 98 per cent contour are marked with blue triangles. The non-JF population is similarly marked by red contours and dots. Histograms show the normalized distributions along each axis. In panel (c), 10 JF galaxies with possibly misattributed hosts are marked in green, as explained in Section 5.2.3. The green marker size shows the ratio of highest local density over all hosts and the density of the assigned host. Arrows point to the location of the JF galaxy on the plane if the high-density host is the true one.

JF at mass of $10^{10} M_{\odot}$ and above, which is seen most clearly in the TNG100 sample – see considerations discussed in Section 5.1.1.

When examining the radial positions of JF galaxies in Fig. 16 we find that the majority of JF galaxies are found within $R_{200,c}$ of their host halo,¹⁰ with the distribution peaking at $\gtrsim 0.5R_{200,c}$ and dropping to low numbers beyond $R_{200,c}$, in contrast to the non-JF distribution that is flatter and declines more gradually. There is a drop in JF numbers in the innermost regions, similar to what was found in Yun19: This is due to the smaller volumes these region represents (there is a similar drop in the non-JF distribution, though not as sharp), but may also be due to the proximity to the central galaxy and other satellites in these regions: Inspectors were explicitly instructed not to classify galaxies as JF if there were other galaxies nearby in the image (Section 2.3.2). Finally, as shown by Yun19, the lack of JF in the cores of groups and clusters is due to the fact that, satellites who reside mostly in the innermost regions of their host, have typically already lost the majority if not all their gas: this both acts against their selection for inspection as well as against the possibility of exhibiting gaseous tails. These results are consistent with observational findings, for example, fig. 5 of Poggianti et al. (2016) and Gullieuszik et al. (2020).

5.2.1 JF galaxies beyond $R_{200,c}$

There are 1168 JF galaxies from our sample that are found beyond $R_{200,c}$ of their host. In the TNG100 sample, the fraction of JF found beyond $R_{200,c}$ is small, but in TNG50 sample there is a substantial number of JF galaxies beyond $R_{200,c}$ and even beyond $2R_{200,c}$. This is in contrast to the non-JF population where the decline with host-centric distance is more gradual, and a non-negligible fraction of satellites can be found even up to $3R_{200,c}$.

We explore this further in Fig. 17 where we show the fraction of both the JF and non-JF populations found beyond $1R_{200,c}$, $1.5R_{200,c}$, and $2R_{200,c}$, for each of the inspected samples (TNG50 and TNG100). The fractions for the non-JF population are similar, with almost half of all satellites found beyond $R_{200,c}$, and 12 per cent (15 per cent) beyond $2R_{200,c}$ in the TNG50 (TNG100) sample.

For the TNG50 JF populations, we see that a *quarter* of all JF reside beyond $R_{200,c}$, while only 13 per cent of TNG100 are found in these regions. Comparing Figs 16(c) and (f), we see that this difference is driven by the differences in satellite stellar masses between the samples: low-mass galaxies are more susceptible to RPS and the chances that they will be JF galaxies in the outer regions of their host is higher. Likewise, in higher mass hosts JF galaxies are more common even in the outer regions. Even beyond $2R_{200,c}$ one can find JF galaxies that comprise several per cent of the JF population, suggesting the presence of ambient gas and environmental effects even at these large distances.

A pertinent question concerning the JF galaxies found beyond $R_{200,c}$ is what percentage have previously been found in the inner regions of a host (‘splashback’) – either their current host or a different one, and whether this may have affected their JF status. An in-depth answer to this issue, exploring different definitions of proximity to a host and how a JF status may be affected by it, is beyond the scope of this paper and merits a dedicated study. In what

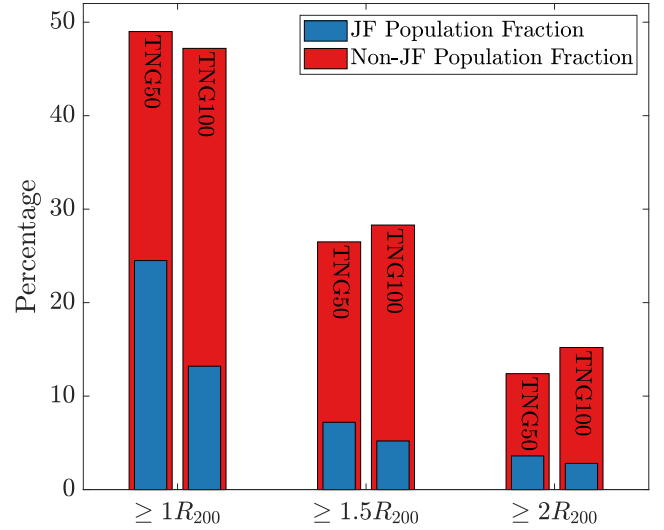


Figure 17. The fraction of the JF and non-JF population found beyond $1R_{200,c}$, $1.5R_{200,c}$, and $2R_{200,c}$, for the TNG50 sample and the TNG100 separately (right and left columns, respectively). JF data are shown in blue while the non-JF data are in red. For the non-JF population, almost half of the satellites are found beyond $R_{200,c}$ (in both samples). Only a quarter of the JF is found beyond $R_{200,c}$ in TNG50, while the fraction of TNG100 JF in that region is only 13 per cent.

follows, we give a preliminary analysis that may serve as a starting point for future endeavors.

To ascertain whether a satellite galaxy has previously been influenced by a host we follow the main progenitor branch of the SUBLINK_GAL merger trees (Rodríguez-Gomez et al. 2015) for all satellites in our sample that are found beyond $R_{200,c}$, both JF and not. For each such outskirts galaxy, we find the time interval between the current time (i.e. at inspection), and the last time it was found within $R_{200,c}$ of a host (either the current host or any other). This time interval should be compared to the sound-crossing time of the satellite galaxy, t_{sc} , which gives the time-scale for propagation of hydrodynamic processes. If the time interval between a the inspection of a JF galaxy and the last instances it was found within a host is large compared to t_{sc} , then it is unlikely the JF features are due to the affect of the host environment accrued when the satellite was within $R_{200,c}$ of that host.

The sound-crossing time for a system of size R is $t_{sc} = R/c_s$, with the sonic velocity of an ideal gas of temperature T given by

$$c_s = \sqrt{\gamma \frac{k_B T}{\mu m_p}}, \quad (6)$$

with an adiabatic index of $\gamma = 5/3$ for a mono-atomic gas. The factor k_B is the Boltzmann constant and μm_p is the average particle mass (m_p is the proton mass). Calculating t_{sc} for a multiphase galactic system including the ISM and CGM can be an arduous affair, but an estimation of this time-scale for the system on the scale of $R_{200,c}$ can be obtained by using the virial temperature of the system

$$T_{vir} = \frac{\mu m_p}{k_B} \frac{G M_{200,c}}{R_{200,c}}, \quad (7)$$

resulting in the following estimate

$$t_{sc} = \frac{R_{200,c}}{c_s} = \left(\gamma \frac{G M_{200,c}}{R_{200,c}^3} \right)^{-1/2} = \left(\frac{4\pi}{3} \gamma 200 G \rho_{crit} \right)^{-1/2}. \quad (8)$$

¹⁰The association of a satellite galaxy with a host and its position with respect to it are determined by the host halo and sub-structure identification methods used in the simulations, namely FoF for the host haloes and SUBFIND for the satellites.

Here, we have used the definition of the virial radius as enclosing a mass whose mean density is 200 times the critical density of the universe.

Estimating the sound-crossing time on the scale of $R_{200,c}$ results in a time-scale that is proportional to the dynamical time of the system and thus has the advantage of being dependent solely on redshift (through the critical density of the universe) and is identical for all systems regardless of mass. For reference, $t_{sc} = 1.1$ Gyr at $z = 0$ and $t_{sc} = 0.63$ Gyr at $z = 1$, for the cosmological model used in the IllustrisTNG simulation suite (Section 2.1).

Of the JF galaxies in our sample found beyond $R_{200,c}$, 9.2 per cent have never previously been within $R_{200,c}$ of any host, compared to 36.6 per cent in the non-JF population. Conversely, 47.9 per cent of the JF galaxies beyond $R_{200,c}$ have been within $R_{200,c}$ of a host within one sound-crossing time, and 72 per cent within $2t_{sc}$. For comparison, among the non-JF outskirt galaxies, these fractions are 26.2 and 46.5 per cent within 1 and 2 sound-crossing times, respectively.

Based on these preliminary results, we conclude that in most JF galaxies found in the halo outskirts, it is likely that the JF status is due, to some extent, to the galaxy having been recently (in terms of sound-crossing times) in the inner regions of a host. However, there is still a non-negligible fraction of JF galaxies where this is not the case, showing that the environmental influence of a host can extend beyond $R_{200,c}$, as shown previously (e.g. Bahé et al. 2013; Cen 2014; Jaffé et al. 2015; Zinger et al. 2018; Ayromlou et al. 2019; Rohr et al. 2023).

5.2.2 JF galaxies in low-mass hosts

The most favourable conditions for the formation of JF galaxies are found in high-mass cluster-sized host, where the higher ambient density and large infall velocities create a strong RPS force. This is clearly evident in our inspected sample: nearly all hosts of mass $10^{14} M_{\odot}$ and above host JF galaxies (Fig. 14), and the highest JF fractions are found in these objects (Fig. 12). Indeed, most observational surveys for JF galaxies have focused on these objects, though JF galaxies have also been observed in low-mass clusters (Poggianti et al. 2017; Gullieuszik et al. 2017), in group-scale hosts (Vulcani et al. 2018, 2021), and even in hosts of $\sim 10^{12} M_{\odot}$ (Poggianti et al. 2016).

However, we find that most JF galaxies are found in group-sized hosts of masses 10^{13} – $10^{14} M_{\odot}$ (see Fig. 16), and furthermore, there is a sizeable population of haloes below the group scale that host JF galaxies. Within our inspected sample, there are ~ 700 hosts of mass below $10^{13} M_{\odot}$ which host roughly 17 per cent of all JF galaxies. Nearly 4 per cent of JF galaxies are found in hosts of mass $10^{12.5} M_{\odot}$ and below. There are even nine hosts of masses of $10^{11.6}$ – $10^{12} M_{\odot}$ with identified JF galaxies.

5.2.3 The case of four JF galaxies found in $\sim 10^{10} M_{\odot}$ hosts

In previous figures, we noted four JF galaxies found in hosts of very low mass, $\sim 10^{10.5} M_{\odot}$. However, the next most massive hosts that contain JF galaxies are a full order of magnitude more massive. In addition, as seen in Fig. 16(e), all four satellites are found relatively far from their hosts, with distance ranging between ~ 1.5 and $\sim 4R_{200,c}$, and all are the only satellite galaxy found in the host within the inspected sample (Fig. 14c). While these may be true JF galaxies within these hosts, other explanations exist.

First, a classification error may falsely report these as JF galaxies. The scores assigned to these objects range between 0.8 and 0.88.

However, a visual inspection by the team experts confirmed that three of the four are clearly JF galaxies, with the fourth also exhibiting some features of JF galaxies. Alternatively, these satellites, while bound to their low-mass hosts, may actually be in the sphere of influence of a more massive host that is responsible for the JF status. We measure the local ambient gas density of the direct hosts of these satellites and compare it to the gas density profiles of other hosts. We examine our entire inspected sample in this manner.

To do so, for each galaxy in the inspected sample, we find its position with respect to *all* haloes within the simulation and then estimate the resulting gas density from *each halo separately* by assuming an NFW profile for the halo density distribution. We also assume the gas density follows the dark matter density, as $\rho_{\text{gas}} = f_{\text{gas}} \rho_{\text{NFW}}(r; M_{200,c}, c_{\text{vir}})$, where r is the distance between the galaxy and the halo centre. The NFW density profile for each host is set by its virial mass parameter, $M_{200,c}$ in this case, and the concentration parameter c_{vir} . The concentration parameter is randomly selected from the $c_{\text{vir}}-M_{200,c}$ relation of Dutton & Macciò (2014). The gas fraction f_{gas} is set as the ratio of the total gas mass and the total dark-matter mass in the halo. For each galaxy, we evaluate all these gas density values and identify the halo with the maximal gas density at the location of the galaxy. We compare it to the gas density of the host halo assigned by the halo-finder. We define the ratio of these two densities as $\eta = \rho_{\text{gas, max}} / \rho_{\text{gas, host}}$.

For the entire inspected sample, we find that less than 1 per cent of all galaxies have $\eta > 1.01$ (a 1 per cent excess). None of the galaxies with $\eta > 1$ are found within $R_{200,c}$ of their assigned host – the closest case is for a satellite found at $\sim 1.4R_{200,c}$. These rare cases of competing influence are only relevant in the outskirts of the hosts. There are only 10 JF galaxies with values of $\eta > 1$ in the entire inspected sample, all from TNG50.

In Fig. 16(e), the location of these ten JF galaxies is marked with green circles. The size of the circle corresponds to the ratio η . The arrows in the figure point to the location on the plane set by the mass of the host exerting the dominant influence. We see that all but one of these 10 galaxies are found at large distances ($> 2.5R_{200,c}$) from their assigned hosts. In nearly all cases, the host exerting the stronger influence is much more massive, and the JF galaxies are actually relatively closer to the host (in terms of the new host $R_{200,c}$). In particular, three of the four JF found in hosts of $\sim 10^{10} M_{\odot}$ are indeed affected by a substantially more massive host ($\gtrsim 10^{13} M_{\odot}$).

6 DISCUSSION

6.1 What are JF tails made of?

As stated previously in Section 2.3, our classification methods are geared towards identifying gas tails that are comprised mainly of material stripped from the satellites' ISM, as is the case for most observational endeavours of these objects. In a companion paper (Rohr et al. 2023, section 3.2), we confirm this by studying the physical properties of the gas tails and show that they are similar to the ISM gas in the main body of the galaxy; furthermore, Göller et al. (2023, section 3.2) find evidence of star formation occurring in the gas tails of the JF galaxies in our sample.

The stripping of the CGM of galaxies by RPS has been considered in various studies (e.g. Kawata & Mulchaey 2008; Bekki 2009; Cen 2014; Jaffé et al. 2015; Zinger et al. 2018; Ayromlou et al. 2019; Wright et al. 2022; Kulier et al. 2023, Rohr et al. 2023b), and it is expected that this would also lead to the formation of filamentary 'tails' of stripped material. However, we argue that to identify these 'JF haloes' in a similar manner one would have to generate a different

Table 3. Selection and results for the test on the viewing angle of the gas maps for the visual inspection of JF galaxies.

Redshift	Total	TNG50		Total	TNG100	
		JF random	JF Opt.		JF random	JF Opt.
$z = 0.5$	1501	94 (6 per cent)	130 (9 per cent)	3064	110 (4 per cent)	177 (6 per cent)
$z = 0$	1417	118 (8 per cent)	161 (11 per cent)	2780	210 (8 per cent)	203 (7 per cent)
Total	2918	212 (7 per cent)	291 (10 per cent)	5844	320 (5 per cent)	380 (7 per cent)

Notes. A subset of all inspected galaxies from both TNG50 and TNG100 at two different redshifts have been used for this comparison: we give here the total numbers of inspected objects as well as the number of JF galaxies identified in the random versus optimal orientations. JF fractions to the inspected samples are given in parenthesis.

set of images, optimized for the lower gas densities of these tails and their spatial extent, and adjust the classification process accordingly. In short, a separate and dedicated classification project would be required to focus on this stage of galaxy evolution.

6.2 Jellyfish galaxies across their evolutionary pathways

In the previous sections, we have focused on *populations* of galaxies selected at various cosmic epochs from the IllustrisTNG simulations. This is formally akin to what is typically possible with observations, with the difference that, within the simulated volumes, galaxies of given epochs are the progenitors or descendants of galaxies at other epochs. Hence, not all among our visually identified JF are unique.

In fact, the sample we identified for visual inspection for the CJF Zooniverse project includes all relevant satellite galaxies from the snapshots of the TNG50 and TNG100 simulations, as described in Section 2.2. In many cases, our inspected sample includes several instances of the same galaxy across a number of different snapshots, which is thus inspected multiple times along its evolutionary path. The sample selection process, and of course the visual classification, are totally agnostic to this. However, scientific results based on our JF scores and analysis need to be interpreted accordingly. In fact, it is also possible to exploit all this by following individual galaxies, and hence their JF score, across their evolutionary pathways.

With the available simulation data, we can link galaxies that have been inspected at multiple cosmic epochs using the SUBLINK.GAL merger trees (Rodríguez-Gómez et al. 2015). Whereas the total sample of inspected satellites amounts to more than 80 000 objects, of these, 5023 and 9052 represent unique galaxy evolutionary tracks in TNG50 and TNG100, respectively. Therefore, on average, each TNG50 (TNG100) galaxy meeting the selection criteria have been inspected ~ 11 (~ 3) times. TNG50 galaxy tracks have been inspected more frequently because every snapshot of this simulation since $z = 0.5$ is included in our inspected sample. We expand on this complementary way to analyse and look at the visually inspected JF in IllustrisTNG in two companion papers (see Rohr et al. 2023, section 2.3; Göller et al. 2023, section 3.5), where we follow satellites across times to quantify the modalities of RPS and their star formation histories. Here, we note that, of the ~ 14 000 unique branches in total inspected in the CJF Zooniverse project, 935 and 922 from TNG50 and TNG100, respectively, are classified as JF at least at one inspected snapshot across their lifetime (adjusted scores).

6.3 Random versus optimized image orientation

The signature feature of a JF galaxy is the asymmetric gas ‘tails’ that trail the main stellar body. As such, identifying a JF galaxy depends on the direction in which one views a JF galaxy. For example, a head-on viewing angle may partially or completely obscure the tails.

To assess this impact, i.e. the projection effects that would also affect any observational survey, we select a sub-sample of objects to show in two orientations: once with a random orientation (the default for the entire inspected sample), and again with an image generated with a viewing angle optimized for identifying tails generated by RPS. For the images in preferred orientations, the gas cell and star particle positions are first rotated about the centre of the galaxy, such that the velocity vector of motion of the galaxy is within the plane of the image, but in a random orientation within the plane. The velocity vector is measured as the bulk peculiar velocity of all particles/cells belonging to the subhalo.¹¹ These images allow us to examine the impact of image orientation on the classifications, by placing any potential gas tails parallel to the image plane, under the assumption (verified in Yun19) that the gas tails are formed in the direction opposite to the direction of motion of the galaxy.

The rest of the classification procedure is the same: The optimally aligned images are shown to inspectors without any special distinction to avoid bias in the classification. A similar comparison in the Yun19 pilot project concludes that as many as 30 per cent of JF galaxies may be missed due to the viewing angle.

For this comparison study, we use all inspected galaxies from both TNG50 and TNG100 in two snapshots: $z = 0, 0.5$. This test sample consists of 8762 objects, of which 532 (6 per cent) are JF based on the randomly oriented images and 671 are JF based on the optimized orientation. The composition of this test sample in terms of simulation and snapshot and the results of the visual classification are shown in Table 3, as well as the number and fraction of JF galaxies based on the random and optimized orientations. Similarly to Yun19, we quantify that as many as 20–30 per cent of JF galaxies may be missed because of an unlucky projection.

In Fig. 18, we show several examples of galaxies that were not identified as JF in one orientation but that, upon changing the viewing angle, received a higher score above the JF threshold. In Figs 18(a)–(c), the randomly oriented image obscures the true extent of the gas tails resulting in images that are rightfully given low scores. The re-orientation of the image brings the gas tails into full view as well as evidence of bow shocks, which further cement a JF identification. In Fig. 18(d), a galaxy with prominent tails is seen with an additional companion, leading to a low score (as requested in the classification instructions). On the other hand, the optimally oriented image shows that the proximity to the other object is due to projection effects and the galaxy is given a much higher score.

However, the re-orientation of the objects does not always lead to better classification. In Figs 18(e)–(h), we showcase four examples of objects classified as JF in the random orientation that received scores

¹¹ We confirmed that using the velocity of the subhalo relative to its host FoF group did not produce a significantly different image and therefore, chose to use the subhalo velocity for simplicity.

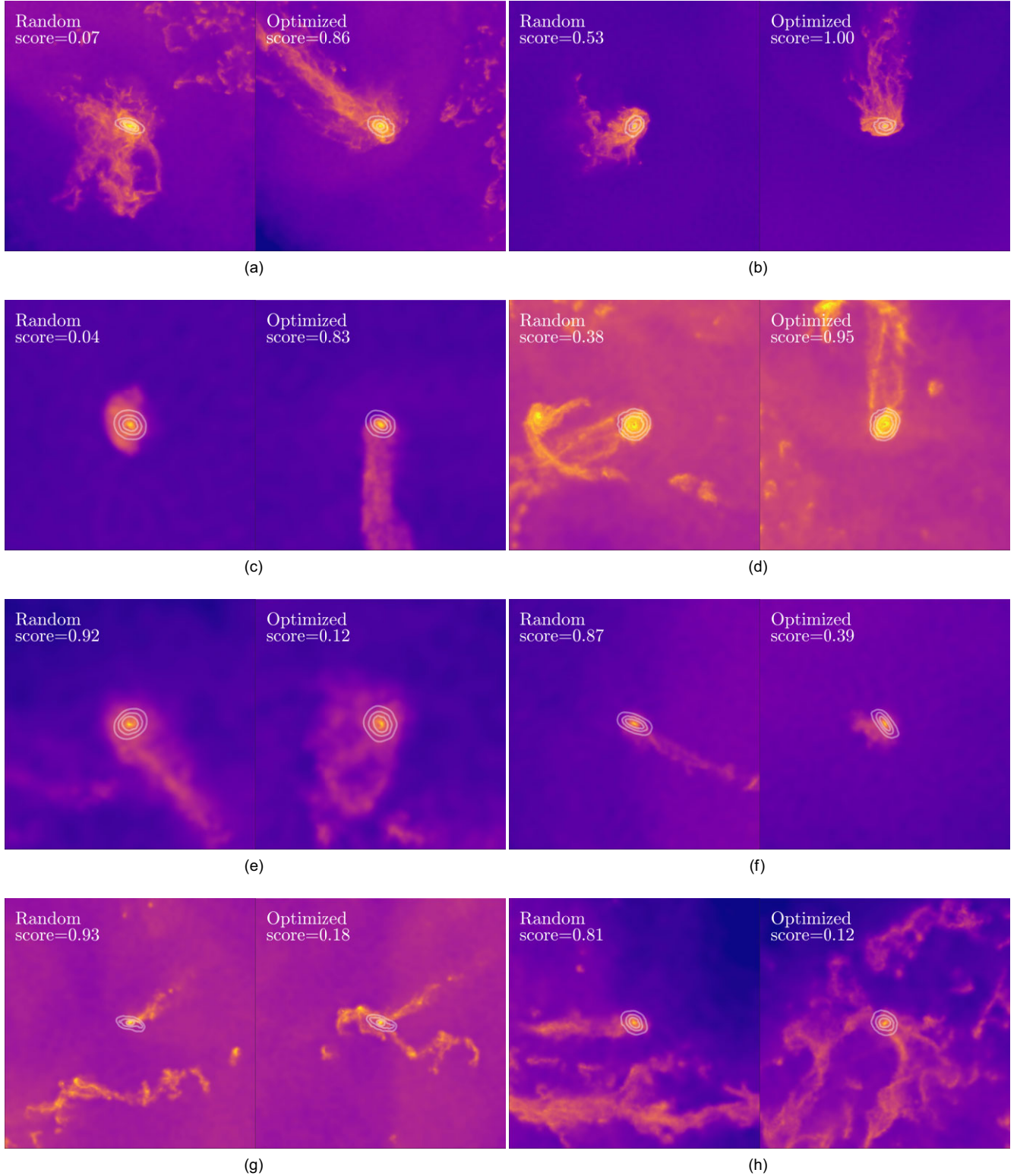


Figure 18. Effect of viewing angle i.e. quantification of projection effects for the identification of JF galaxies in gas maps. We show pairs of images for the same IllustrisTNG objects where the optimized viewing angle resulted in a higher score and JF classification (Figs 18a–d), and examples of cases where the *random* viewing angle resulted in a higher score and JF classification (Figs 18e–h). The latter cases are rarer than the former ones.

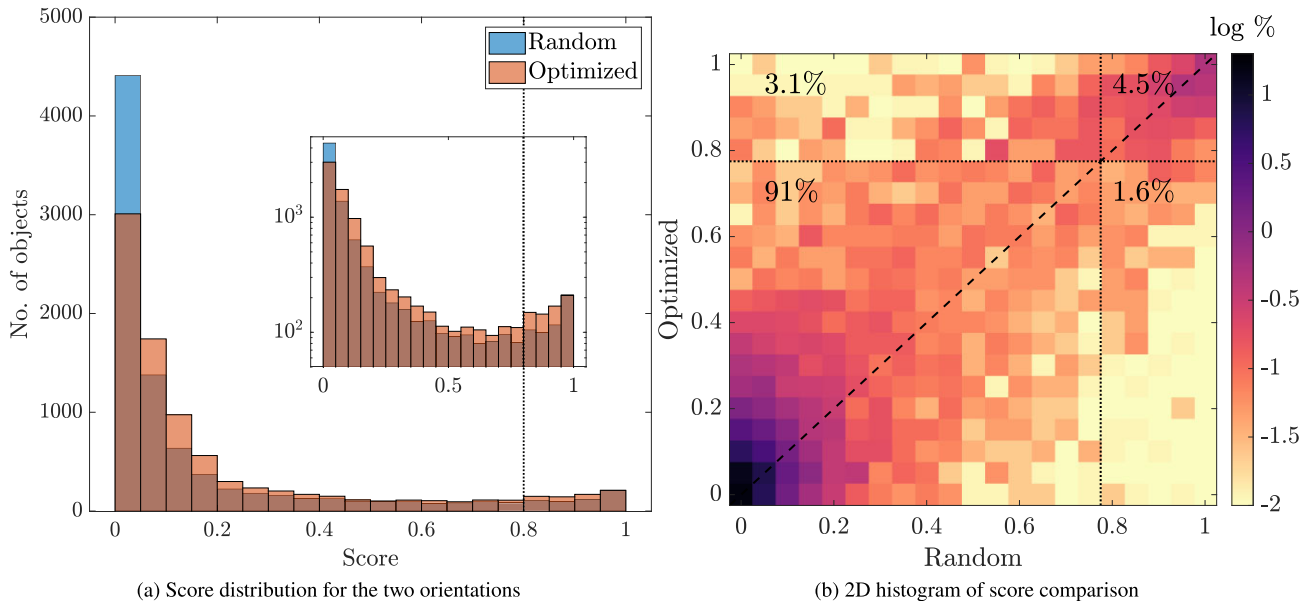


Figure 19. Comparison of JF scores (adjusted) between images shown in random versus optimized orientations. The distributions of scores for the random (blue histogram) and optimized (red histogram) projections are shown in panel (a), with the inset showing the histograms on a log scale. The vertical dotted lines denote our fiducial JF threshold score of 0.8. Panel (b) shows the 2D histogram comparing the classifications with different viewing angles, with the axes reporting the scores based on the randomly oriented (x-axis) and optimized (y-axis) images. The colours are the percentage of the population in each pixel. The horizontal and vertical dotted lines mark the JF threshold score, and the percentage of the population in each quadrant is also provided.

below the JF threshold in the optimal orientation (out of 136 such cases). In Figs 18(e) and (f), we see that the optimal viewing angle actually makes the tails *less* visible, while in Figs 18(g) and (h) the re-orientation leads to projection effects which mask the appearance of the gas tails. These cases are, however, twice as rare (within this testing sample) than the ones whereby the optimal projection returns higher JF scores, as we quantify next.

Fig. 19(a) quantifies the distribution of scores these objects received, both in the random and optimized orientations. As a result of the improved viewing angle, many ($\gtrsim 1000$) objects of score zeros now have slightly higher scores. In the JF region (≥ 0.8), there is an increase of several hundred objects. However, the overall shape of the distribution does not change dramatically.

A more detailed view can be found in Fig. 19(b), where we show a 2D histogram contrasting the scores based on the random and optimized orientations. Horizontal and vertical lines mark the JF threshold score and divide the plane into quadrants: two quadrants (bottom-left and top-right) in which the objects do not change their JF classification despite a change in score and two quadrants (top-left and bottom-right) in which the change in score led to a different classification – non-JF to JF and vice versa. The percentage of the sample in each quadrant is shown as well.

Over half of the objects in the sample increase in score (55 per cent), about 30 per cent decrease in score and 15 per cent do not change at all. There is a net increase of 26 per cent in the number of JF galaxies identified by the optimized orientation. Within the inspected test sample, there are 136 objects that are identified as JF galaxies in the random orientation but not in the optimized orientation. Most of these objects have scores that are close to the JF threshold in the random orientation.

If we consider the optimized classifications as the true JF population and the random classifications as a representation of what an observational survey may produce, we see that the viewing angle may lead to missing 40 per cent of the true JF galaxies (objects in

the top-left quadrant). In addition, roughly 26 per cent of the objects thought to be JF are in fact not (bottom-right quadrant), resulting in the *net* of 26 per cent. However, upon inspection of the objects in this quadrant, a large number of them are indeed JF, with the optimized projection either reducing the visibility of the tails (as shown in Fig. 18), or presenting a very similar image as the random projections but with a reduced score. If the objects in the lower right quadrant are still considered JF, then adding the JF objects from the new projection results in a 50 per cent increase in JF numbers.

To summarize, roughly a third of all JF galaxies may go undetected because of viewing angle-related issues, in agreement with Yun19.

6.3.1 Different scores for (nearly) identical images

Inspection of the objects in the top-left and bottom-right quadrants of Fig. 19(b) yield a set of ~ 35 galaxies (out of 329) whose images in the two orientations are so similar as to be practically identical (though some are rotated in the image plane), but which received significantly different scores. We show six examples of these objects in Fig. 20. The score difference for all such objects we found is on average ~ 0.2 in the 0–1 score range. We discuss the implication of this finding in Section 6.4.

6.4 Overall validity of the JF identification method

The results in this paper and future ones that will be based on the CJF Zooniverse project rely on the classification method used to identify JF galaxies in our inspected sample. In this section, we discuss the advantages and drawbacks of our method and suggest improvements for the future.

Our classification method has its drawbacks, some common to all citizen-science projects and some particular to the field of JF galaxies. The primary issue is that the majority of the classifications are carried out by volunteers who have little to no background knowledge of

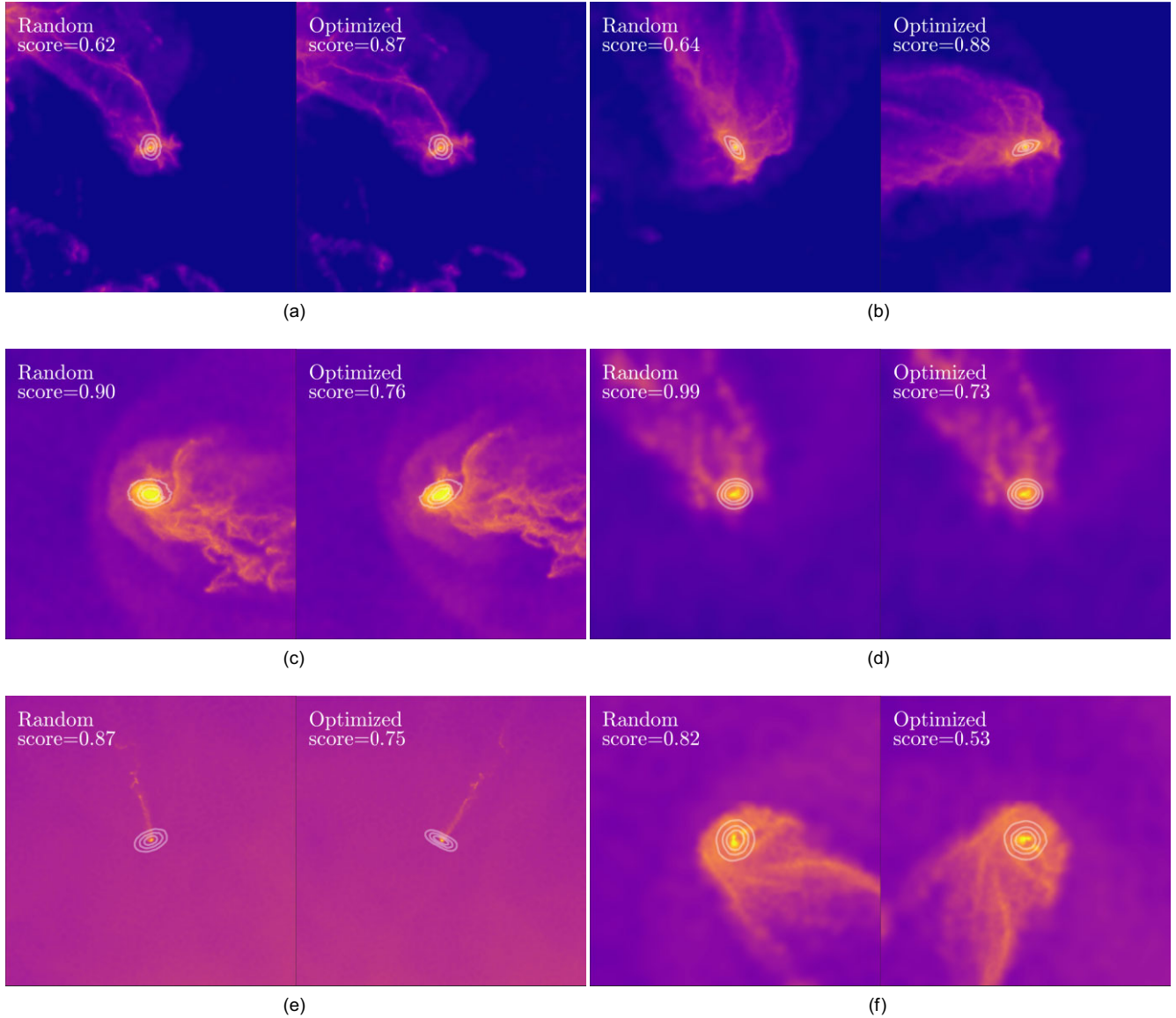


Figure 20. Examples of IllustrisTNG galaxies where the optimized viewing angle resulted in images that are very similar to those in random orientation, but with a difference in score (see the text for details).

what a galaxy is, let alone a JF galaxy. We argue that the nature of the task makes this inexperience a minor handicap: a galaxy that is clearly not a JF and one that clearly is, is relatively easy to identify. But in cases with some ambiguity, such as a messy background, multiple tails, or extended gas haloes surrounding the stellar body, the choice is not always clear even to experts. In such cases, prior experience and expert knowledge is important to find the secondary features that help classifications: the existence of bow-shocks, signatures of outflows, tail continuity, etc.

The tutorial and training module provided to volunteers partially address this lack of knowledge and experience, and provide a common level of basic knowledge required for the task. The tutorial includes several examples of the galaxies of interest, and vice versa. However, the diversity and richness in the images make it impossible to show examples of all the different features volunteers may encounter while performing the task.

An additional challenge is the subjectivity and bias inherent in any human-based classification. A citizen-science task must be simple

and straightforward. Our task-question: ‘Do you think that the galaxy looks like a JF?’ is by its nature asking the inspector for a subjective answer. As the examples shown in Section 6.3.1 exemplify, two nearly identical images may garner different scores from different inspectors. Even when galaxies appear identical, a built-in bias for tail direction may increase or decrease the likelihood of a ‘yes’ answer to our question. The effect of human bias in citizen-science has also been observed when selecting chirality and spiral structure in the Galaxy Zoo project (Land et al. 2008; Hayes, Davis & Silva 2017).

One way to alleviate these issues is to expand the number of inspectors per image who determine the final score. If enough people cast their votes, personal inexperience, errors in judgement and biases can be compensated for. The common wisdom is that the larger the number of votes, the more dependable the result is (as demonstrated in Section 4.2). However, this choice must also take into account practical considerations: requiring too many votes per image may extend the duration of the classification phase. In addition, if a project appears to be advancing very slowly, volunteers may lose interest

and abandon the project altogether. This choice must be made while carefully balancing scientific necessity and resource-management.¹² While our choice of 20 independent votes per image may have led to some cases of mis-classification, we believe it achieves the dual goals of providing a mostly reliable classification within a reasonable time-frame.

As a result, we have carried out post-processing steps to increase our confidence in the results. First, the classification comparison between the CJF project and the pilot projects (Section 3.2) yields good agreement between volunteers and expert classifications. Secondly, the inspector-weighting scheme marginalizes over inspectors with little experience or bad track-records, and adds weight to experts (Section 3.3 and Appendix A). Thirdly, the high threshold for identifying JF galaxies was chosen to achieve a very pure JF sample, even at the expense of leaving some JF galaxies out.

Thanks to these steps, we are confident that our final sample is indeed comprised of JF galaxies and can be relied upon to produce accurate information about the JF population.

Additional measures may further improve the quality of citizen-science classifications in the future. For example, carrying out expert classifications on a random sampling of objects. This allows the team to gauge how well the classification aligns with an expert opinion while adding the expert vote to the final tally. This may be particularly useful for borderline scores just below the threshold, where there are surely JF galaxies that have been missed, as is evident in Figs 10 and B4. Studies into the bias and overall quality of classifications may also be useful, for example, presenting the same galaxy from various viewing angles to further explore the dependence of classification on viewing angle, exploring image-enhancing techniques such as background subtraction and showing the same *image* multiple times rotated in different directions to detect possible bias.

7 SUMMARY AND CONCLUSIONS

In this work, we have presented the results of the CJF Zooniverse project: a citizen science effort to identify and classify JF satellite galaxies in the IllustrisTNG simulations. Over 6000 volunteers classified nearly 90 000 images of over 80 000 satellite galaxies.

The inspected galaxy sample is unprecedented in terms of its size, the satellite stellar mass and host mass ranges, and the redshift coverage: it contains all satellite galaxies in the TNG50 and TNG100 simulations above a satellite stellar mass of $10^{8.3}$ and $10^{9.5} M_{\odot}$, respectively, and with some gas. There is no explicit minimal or maximal mass cut for the hosts, which range from $10^{10.4}$ up to $10^{14.6} M_{\odot}$ and include over 2000 group-sized and over 100 cluster-sized systems (TNG100). Ten snapshots of the TNG100 box were included between $z = 2$ and 0 (intervals of ~ 1 Gyr). For TNG50, four snapshots between $z = 2$ and 0.5 were examined, identical in redshifts to the TNG100 snapshots, and in addition, *all* available snapshots between $z = 0.5$ and 0 were inspected allowing us to examine the state of a galaxy (JF or not) over time intervals of ~ 150 Myr. Some of the images hence include the same galaxies across several snapshots (Sections 2.2 and 6.2).

The aim of our classification process was a sample of simulated galaxies that could be confidently called ‘JF’ galaxies. We preferred a pure sample rather than a complete one and this choice set the

guidance we provided the volunteers with, the algorithm chosen to assess the inspectors, and the threshold chosen to identify JF galaxies.

Images of the gas density field for all galaxies in the inspected sample were created along a random orientation (Section 2.3.1), and presented on the Zooniverse platform to at least 20 volunteers with a simple yes/no question: ‘Does this galaxy look like a JF?’ (see Section 2.3.2 for details), resulting in a set of raw scores ranging between 0 and 20. To validate the process, a comparison against the classifications by a team of experts on two previous smaller visual-inspection campaigns was carried out: the Yun19 (based on TNG100) sample and a pilot project of TNG50 images. The agreement with the TNG50 pilot is excellent, whereas the agreement with the Yun19 sample is slightly lower, likely due to differences in the types of images used for the classification (Section 3.2). In any case, both comparisons lend support to the usage of the classification by non-experts, with a high enough number of classifications per object, for the purpose of identifying JF galaxies in gas maps.

Furthermore, we have implemented an inspector-weighting algorithm to minimize the impact of inexperienced volunteers and those that consistently voted against the consensus, and added weight to experts inspectors, i.e. research team members (Section 3.3 and Appendix A). This results in a set of *adjusted* scores that we recommend for usage and that range between 0 (definitely *not* a JF) and 1 (most definitely a JF), with the threshold value for a JF galaxy set to be 0.8 (Section 4.1).

An additional set of images, produced along an optimized orientation for tail identification for a subset of the TNG50 and TNG100 galaxies (at $z = 0.5$ and 0) allowed us to carry out a preliminary analysis of the effect of viewing angle i.e. projection effects. By a conservative estimation, a third of all JF galaxies are mis-classified (i.e. missed) due to the viewing angle (Section 6.3). A set of galaxies in which the two orientations give practically identical images, but which receive different scores (by ~ 0.2 , on average, in the 0–1 score range), showed the inherent bias issues in citizen-science projects.

7.1 Guidelines for the use of the CJF data set

With this publication, we make the full data set of JF scores generated by the CJF project publicly available via the [IllustrisTNG website](#), for future usage and reference. The data set includes separate galaxy catalogues for each snapshot in the TNG50 and in the TNG100 simulations. Each catalogue lists the raw scores for each galaxy, as well as the adjusted score (after implementing the inspector weighting), both based on the map inspected with random orientations. The total weight of the combined inspectors and the number of experts that viewed each galaxy image is also included in the data set. In addition, the scores for the optimized orientation images are also available. In this paper, we have motivated our recommended score threshold of 0.8 (in the 0–1 range of adjusted scores) above which a galaxy can be called ‘JF’ and we have inspected the basic demographics of so-defined JF galaxies from the IllustrisTNG simulations. In Section 4.2, we suggest a statistical framework that can be used to gauge the confidence level associated with this or an alternative choice of threshold score.

7.2 Demographics of the JF population

Our JF sample contains 5307 galaxies from the TNG50 and TNG100 simulations, comprising ~ 7 per cent of the *inspected* satellite population. As a reminder, the latter does not include the entire populations of simulated satellites above a certain stellar mass but includes only

¹²The main resource is the volunteer pool, but maintaining a citizen-science project also requires significant time and effort from the research teams in preparing the project, interacting with volunteers via message boards, volunteer recruitment efforts, and so on.

those with at least some gravitationally bound gas – a fundamental requirement to witness RPS in action.

General demographics: Over the entire population of inspected satellites, JF galaxies are more frequent in higher mass hosts and for smaller satellite stellar masses. Among the lower mass inspected satellites ($\lesssim 10^{10} M_{\odot}$) in cluster-sized systems, about 40 per cent are JF. Within the inspected satellite population in group-sized system JF galaxies comprise ~ 25 per cent. The JF galaxies are more abundant at low masses than their non-JF counterparts, with very few JF of stellar mass above $10^{10} M_{\odot}$. This is due in part to the increased gravitational binding force, but also may be affected by the onset of the kinetic feedback mode in the IllustrisTNG physical model (at $\sim 10^{10.5} M_{\odot}$), that may deplete the gas from the satellite.

The distribution of host masses for JF galaxies is skewed to higher masses compared to the general satellite population. Most haloes above $10^{13.5} M_{\odot}$ contain JF galaxies and nearly every cluster-sized system, $10^{14} M_{\odot}$ and above, has a JF population, with JF fractions of 10–40 per cent (Section 5.1, Figs 12, 14, and 16).

JF in low-mass hosts: Over half of all JF are found in groups and clusters. While JF fractions increase with halo mass, the peak of the distribution of JF in haloes mass is at the group scale, $10^{13.5} M_{\odot}$, due to the much higher number of hosts with lower masses. However, nearly 1000 JF galaxies reside in hosts of mass less than $10^{13} M_{\odot}$, of which 200 are in hosts of mass $10^{12.5} M_{\odot}$ and below. The least massive hosts that host JF galaxies are $10^{11.5} M_{\odot}$ in TNG50 and $10^{12.5} M_{\odot}$ in TNG100 (Section 5.2.2, Figs 14 and 16).

JF beyond $R_{200,c}$: While most JF are concentrated within $R_{200,c}$ (compared to the general satellite population), a full *quarter* of all JF may be found beyond it. A few per cent may even be found beyond $2R_{200,c}$. These JF in the outskirts are more likely to be of lower masses and to reside around higher mass hosts. While most of these JF galaxies may have been found within $R_{200,c}$ of a host in their recent past, a non-negligible fraction have most likely been shaped by the environmental influence of their current host that extends beyond the $R_{200,c}$ scale (Section 5.2.1, Figs 16 and 17).

JF at high redshifts: Our inspected sample extends up to $z = 2$, and we find JF galaxies at all redshifts. At redshifts of $z = 1.5$ – 2 , they are found in small groups and protoclusters (10^{12} – $10^{14} M_{\odot}$) where they comprise ~ 10 per cent of the satellite population in these systems (there are ~ 40 such objects that represent 3 per cent of all haloes in this epoch and in this mass range). JF fractions as a function of host mass show little redshift dependence, while the JF fractions in satellite stellar mass bins grow with decreasing redshift (Section 5.1, Figs 13, 14c and d).

The results presented above showcase the enormous potential this data set represents, thanks to the combination of the large volumes simulated within the IllustrisTNG project with the capacity of citizen-science projects such as those configurable on Zooniverse. The ability to study such a large sample of JF galaxies, which span a large range of properties, environments, and cosmic epochs, will enable us to address many open questions about galaxy evolution in groups and clusters and in particular about this important phase in the life-cycle of satellite galaxies. In two companion papers, we study JF galaxies along their evolutionary paths and learn when and where they lose their cold gas (Rohr et al. 2023) and quantify their star formation properties and histories (see Göller et al. 2023).

The sheer size of the inspected and JF samples provided here will also allow us to expand the identification to even larger data sets of simulated and observed galaxies (e.g. to the TNG300 simulation and others), via machine learning. The data set presented in this work can constitute a valuable resource both for training convolutional neural networks – preliminary tests in this direction with the Yun19 data

showed much promise – as well as for comparison to the results of unsupervised or self-supervised classification schemes.

Finally, whereas the primary goal of the CJF Zooniverse project was a scientific one, offering a (surprisingly) fast way of carrying out visual classification of a large number of images, an additional, and in our opinion highly valuable, goal was achieved: the exposure of the general public to the research of galaxy evolution in general and the study of JF galaxies in particular. This was done through the educational material on the website, interaction on the message boards, and volunteer recruitment efforts (press releases, public talks and others). We hope this may lead, in some small way, to increasing science literacy and to inspiring scientific study, especially in younger audience.

ACKNOWLEDGEMENTS

This publication uses data generated via the [Zooniverse.org](https://zooniverse.org) platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. We wish to extend our thanks to the team at Zooniverse with their advice and assistance in building and running this project. We also thank the thousands of volunteers who invested their time and effort to assist us in this project. We thank the referee for their insightful comments which improved the paper. ER is a fellow of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD). DN acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1). GDJ acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 818085 GMGalaxies. The authors also thank Martina Donnari and Allison Merritt for their participation in the pilot classification project that we performed to verify the function of the ‘Cosmological Jellyfish’ Zooniverse website and that we have used to assess its consistency. EZ thanks Yehuda Hoffman and Yuval Birnboim for their fruitful advice and discussions.

DATA AVAILABILITY

All the simulation data of the IllustrisTNG simulations, and hence of TNG100 and TNG50 used here, are publicly available and accessible at <https://www.tng-project.org/data> (Nelson et al. 2019a). With this paper, we release the scores of the fiducial JF classification justified, produced and analysed here, so that interested researchers and citizen scientists can make further usage of our simulation data and of the visual classification.

The scripts used for the data analysis and figures can be found here: https://github.com/eladzing/matlab_scripts.git. Please contact the corresponding author for assistance at elad.zinger@mail.huji.ac.il.

REFERENCES

- Ayromlou M., Nelson D., Yates R. M., Kauffmann G., White S. D. M., 2019, *MNRAS*, 487, 4313
- Bahé Y. M., McCarthy I. G., Balogh M. L., Font A. S., 2013, *MNRAS*, 430, 3017
- Bekki K., 2009, *MNRAS*, 399, 2221
- Benneyan J. C., Borgman A. D., 2004, Proc. IIE Annual Conference. Institute of Industrial and Systems Engineers (IIE), Norcross, United States, p. 1
- Borne K. D., Zooniverse Team, 2011, AGU, 2011, ED23C

- Borne K. D., Fortson L., Gay P., Lintott C., Raddick M. J., Wallin J., 2009, AGU, 2009, ED51C
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Gavazzi G., Lagos C., Saintonge A., 2014, *A&A*, 564, A67
- Boselli A. et al., 2016, *A&A*, 587, A68
- Boselli A. et al., 2018, *A&A*, 620, A164
- Boselli A. et al., 2019, *A&A*, 631, A114
- Boselli A., Fossati M., Sun M., 2022, *A&AR*, 30, 3
- Cardamone C. et al., 2009, *MNRAS*, 399, 1191
- Catinella B. et al., 2013, *MNRAS*, 436, 34
- Cen R., 2014, *ApJ*, 781, 38
- Chung A., Gorkom J. H. v., Kenney J. D. P., Crowl H., Vollmer B., 2009, *AJ*, 138, 1741
- Cortese L., Catinella B., Smith R., 2021, *PASA*, 38, e035
- Cowie L. L., Songaila A., 1977, *Nature*, 266, 501
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Dressler A., 1980, *ApJ*, 236, 351
- Dutton A. A., Macciò A. V., 2014, *MNRAS*, 441, 3359
- Ebeling H., Stephenson L. N., Edge A. C., 2014, *ApJ*, 781, L40
- Engler C., Pillepich A., Joshi G. D., Pasquali A., Nelson D., Grebel E. K., 2023, *MNRAS*, 522, 5946
- Fumagalli M., Krumholz M. R., Prochaska J. X., Gavazzi G., Boselli A., 2009, *ApJ*, 697, 1811
- Gavazzi G., Perola G. C., 1978, *A&A*, 66, 407
- Gavazzi G., Boselli A., Driel W. v., O’Neil K., 2005, *A&A*, 429, 439
- Gavazzi G., Consolandi G., Gutierrez M. L., Boselli A., Yoshida M., 2018, *A&A*, 618, A130
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Gnedin O. Y., 2003a, *ApJ*, 582, 141
- Gnedin O. Y., 2003b, *ApJ*, 589, 752
- Göller J., Joshi G. D., Rohr E., Zinger E., Pillepich A., 2023, *MNRAS*, 525, 3551
- Gómez P. L. et al., 2003, *ApJ*, 584, 210
- Gullieuszik M. et al., 2017, *ApJ*, 846, 27
- Gullieuszik M. et al., 2020, *ApJ*, 899, 13
- Gunn J. E., Gott III J. R., 1972, *ApJ*, 176, 1
- Hayes W. B., Davis D., Silva P., 2017, *MNRAS*, 466, 3928
- Healy J. et al., 2021, *A&A*, 650, A76
- Jáchym P., Combes F., Cortese L., Sun M., Kenney J. D. P., 2014, *ApJ*, 792, 11
- Jáchym P. et al., 2017, *ApJ*, 839, 114
- Jaffé Y. L., Smith R., Candlish G. N., Poggianti B. M., Sheen Y.-K., Verheijen M. A. W., 2015, *MNRAS*, 448, 1715
- Joshi G. D., Pillepich A., Nelson D., Marinacci F., Springel V., Rodríguez-Gómez V., Vogelsberger M., Hernquist L., 2020, *MNRAS*, 496, 2673
- Kawata D., Mulchaey J. S., 2008, *ApJ*, 672, L103
- Kenney J. D. P., Gorkom J. H. v., Vollmer B., 2004, *AJ*, 127, 3361
- Kulier A. et al., 2023, *ApJ*, 954, 177
- Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, *MNRAS*, 481, 3573
- Land K. et al., 2008, *MNRAS*, 388, 1686
- Lewis I. et al., 2002, *MNRAS*, 334, 673
- Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
- Lintott C. J. et al., 2009, *MNRAS*, 399, 129
- Machacek M., Jones C., Forman W. R., Nulsen P., 2006, *ApJ*, 644, 155
- Marinacci F. et al., 2018, *MNRAS*, 480, 5113
- McPartland C., Ebeling H., Roediger E., Blumenthal K., 2016, *MNRAS*, 455, 2994
- Miley G. K., Perola G. C., Kruit P. C. V. D., Laan H. V. D., 1972, *Nature*, 237, 269
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, 379, 613
- Müller A. et al., 2021, *Nat. Astron.*, 5, 159
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nelson D. et al., 2019a, *Comput. Astrophys. Cosmol.*, 6, 2
- Nelson D. et al., 2019b, *MNRAS*, 490, 3234
- Nulsen P. E. J., 1982, *MNRAS*, 198, 1007
- Pakmor R., Springel V., 2013, *MNRAS*, 432, 176
- Peluso G. et al., 2022, *ApJ*, 927, 130
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Pillepich A. et al., 2019, *MNRAS*, 490, 3196
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Poggianti B. M. et al., 2016, *AJ*, 151, 78
- Poggianti B. M. et al., 2017, *ApJ*, 844, 48
- Ricarte A., Tremmel M., Natarajan P., Quinn T., 2020, *ApJ*, 895, L8
- Roberts I. D., van Weeren R. J., McGee S. L., Botteon A., Ignesti A., Rottgering H. J. A., 2021a, *A&A*, 652, 14
- Roberts I. D. et al., 2021b, *A&A*, 650, A111
- Rodríguez-Gómez V. et al., 2015, *MNRAS*, 449, 49
- Roediger E., 2009, *Astron. Nachr.*, 330, 888
- Roediger E., Brüggen M., 2007, *MNRAS*, 380, 1399
- Roediger E., Brüggen M., 2008, *MNRAS*, 388, 465
- Rohr E., Pillepich A., Nelson D., Zinger E., Joshi G. D., Ayromlou M., 2023a, *MNRAS*, 524, 3502
- Rohr E., Pillepich A., Nelson D., Ayromlou M., Zinger E., 2023b, preprint (arXiv:2311.06337)
- Roman-Oliveira F. V., Chies-Santos A. L., Rodríguez del Pino B., Aragón-Salamanca A., Gray M. E., Bamford S. P., 2019, *MNRAS*, 484, 892
- Ruszkowski M., Brüggen M., Lee D., Shin M.-S., 2014, *ApJ*, 784, 75
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Shostak G. S., Hummel E., Shaver P. A., van der Hulst J. M., van der Kruit P. C., 1982, *A&A*, 115, 293
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
- Smith A. M. et al., 2011, *MNRAS*, 412, 1309
- Smith R. J. et al., 2010, *MNRAS*, 408, 1417
- Smith R. et al., 2022, *ApJ*, 934, 86
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *MNRAS*, 391, 481
- Springel V., 2010, *MNRAS*, 401, 791
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Sun M., Jones C., Forman W., Nulsen P. E. J., Donahue M., Voit G. M., 2006, *ApJ*, 637, L81
- Terrazas B. A. et al., 2020, *MNRAS*, 493, 1888
- Tonnesen S., Bryan G. L., 2009, *ApJ*, 694, 789
- Tonnesen S., Stone J., 2014, *ApJ*, 795, 148
- Tremmel M. et al., 2019, *MNRAS*, 483, 3336
- Verdugo C., Combes F., Dasyra K., Salomé P., Braine J., 2015, *A&A*, 582, A6
- Villalobos A., De Lucia G., Murante G., 2014, *MNRAS*, 444, 313
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Vulcani B. et al., 2018, *MNRAS*, 480, 3152
- Vulcani B. et al., 2021, *ApJ*, 914, 27
- Weinberger R. et al., 2017, *MNRAS*, 465, 3291
- Wood R. A. et al., 2017, *ApJ*, 847, 79
- Wright R. J., Lagos C. d. P., Power C., Stevens A. R. H., Cortese L., Poulton R. J. J., 2022, *MNRAS*, 516, 2891
- Yun K. et al., 2019, *MNRAS*, 483, 1042 (Yun19)
- Zinger E., Dekel A., Kravtsov A. V., Nagai D., 2018, *MNRAS*, 475, 3654
- Zinger E. et al., 2020, *MNRAS*, 499, 768

APPENDIX A: INSPECTOR RANKING

The following describes the procedure for generating weights for all inspectors in the CJF project. Inspector identification is based on the user name assigned to inspectors who are logged into the platform. If an inspector performs classifications while not logged on, a random identifier is assigned that signifies that the user is not logged on.

For each inspector, we assign a set of weights based on different criteria, with the final weight determined by a combination of these, as detailed below:

Table A1. Object weighting for the up-/down-voter ranking.

Object score (low end)	Object score (high end)	Object weight
0.05 and below	0.95 and above	5
0.05–0.1	0.9–0.95	4
0.1–0.15	0.85–0.9	3
0.15–0.2	0.8–0.85	2
0.2–0.25	0.75–0.8	1
Otherwise		0

(i) Inexperienced inspectors – Inspectors that have classified fewer than 10 images are all given a weight of 0.5. As can be seen in Fig. 5, this applies to several hundreds of inspectors.

(ii) Expert inspectors – Members of our research group are all deemed experts and given a weight of 5. Classifications from the previous classification projects, the TNG50 Pilot and the Yun19 projects, are also treated as being carried out by experts.

(iii) Performance on high-score objects – We assess how each inspector performs compared to their peers when classifying high-score objects. In this way, we are giving more weight to inspectors who have shown they can correctly identify a JF when presented with one. The score of an object is the fraction of ‘yes’ votes an object receives in the initial tally. High-score objects are defined as objects that received a score greater than 0.5, and the higher the score of the object, the more important it is in setting the inspector ranking. The inspector weight is defined as the weighted mean of all the ‘yes’ (1) votes of a given inspector over all high-score objects

$$W_{\text{high}} = \frac{\sum_{\text{score} \geq 0.5} w_o v}{\sum_{\text{score} \geq 0.5} w_o}. \quad (\text{A1})$$

With $w_o = 2 \times \text{score} - 1$ being the object weight and v being the vote given for each object (either 0 or 1).

(iv) Identifying up/down-voters – We identify up-voters as inspectors who consistently mis-identify JF galaxies, voting ‘no’ on objects that received predominately ‘yes’ votes (down-voters) and vice versa (up-voters). We assign a rank to each inspector to appraise whether or not they are an up-/down-voter by finding the weighted mean of votes cast against the consensus on objects in the low/high end of the initial tally. The higher the rank, the more likely an inspector is an up-/down-voter.

The ranking of down-voters is calculated on objects of initial score 0.75 and above:

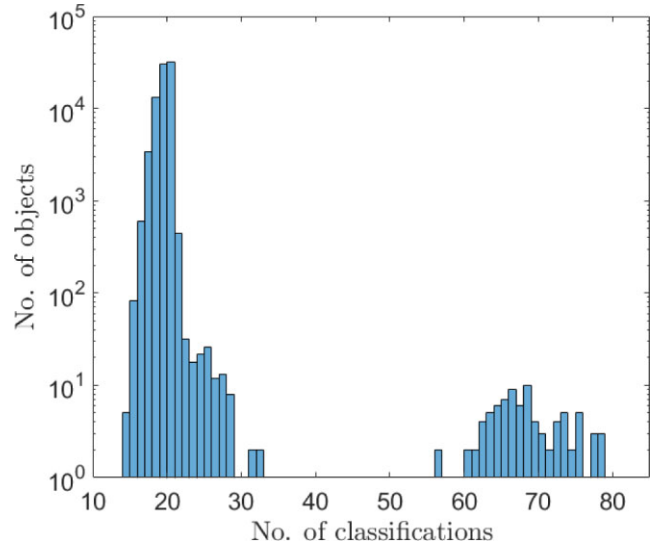
$$R_{\text{down}} = \frac{\sum_{\text{score} \geq 0.75} w_d (1 - v)}{\sum_{\text{score} \geq 0.75} w_d}, \quad (\text{A2})$$

with the object-weighting set by Table A1. Similarly, the up-vote ranking is set by

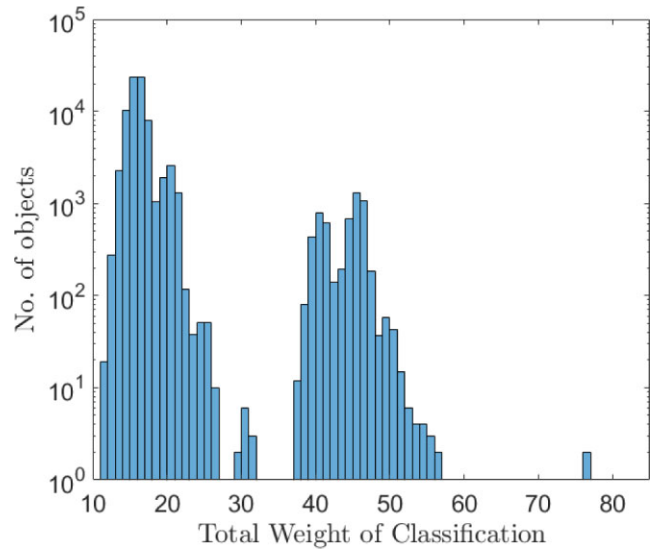
$$R_{\text{up}} = \frac{\sum_{\text{score} \leq 0.25} w_d v}{\sum_{\text{score} \leq 0.25} w_d}, \quad (\text{A3})$$

An inspector that has voted against the consensus at least three times on objects of scores 0.75 and above, and has a down-vote ranking of at least 0.5 is designated a down-voter and given an inspector weight of 0, i.e. all of their classifications are ignored. In a similar fashion, an inspector that has voted against the consensus at least three times on objects of scores 0.25 and below, and has a up-vote ranking of at least 0.5 is designated an up-voter, and given an inspector weight of 0.

In total, we identified 125 down-voters and 213 up-voters.



(a) Classification Histogram



(b) Weight Histogram

Figure A1. The number of classifications for the objects in the inspected sample and the total weight of the inspectors.

The overall inspector weight is set to be W_{high} , unless the inspector is either inexperienced (weight of 0.5) or an expert (weight of 5), or if they have been identified as an up-voter or down-voter and are removed completely from tally (weight of 0).

In Fig. A1, we show the histogram for the number of classifications for the objects in the inspected sample as well as the histogram of the sum of the inspector weights. Although there are now objects that were classified by fewer than 20 inspectors, all objects were classified by at least 13 inspectors.

APPENDIX B: SAMPLE OF GALAXY IMAGES IN SCORE BINS

In Figs B1–B5, we show several examples of galaxies in five bins of JF scores ranging from 0–0.2 up to 0.8–1.

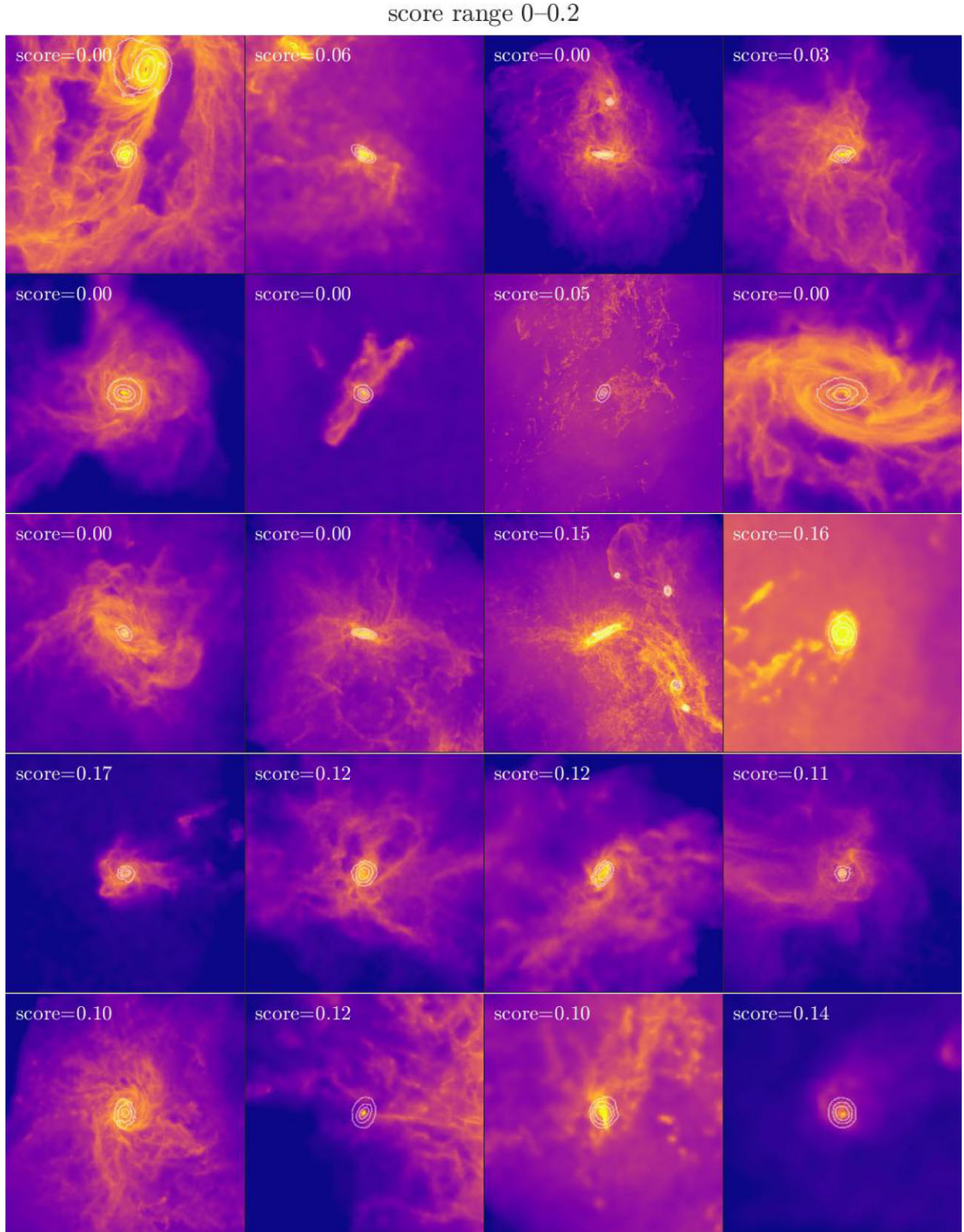
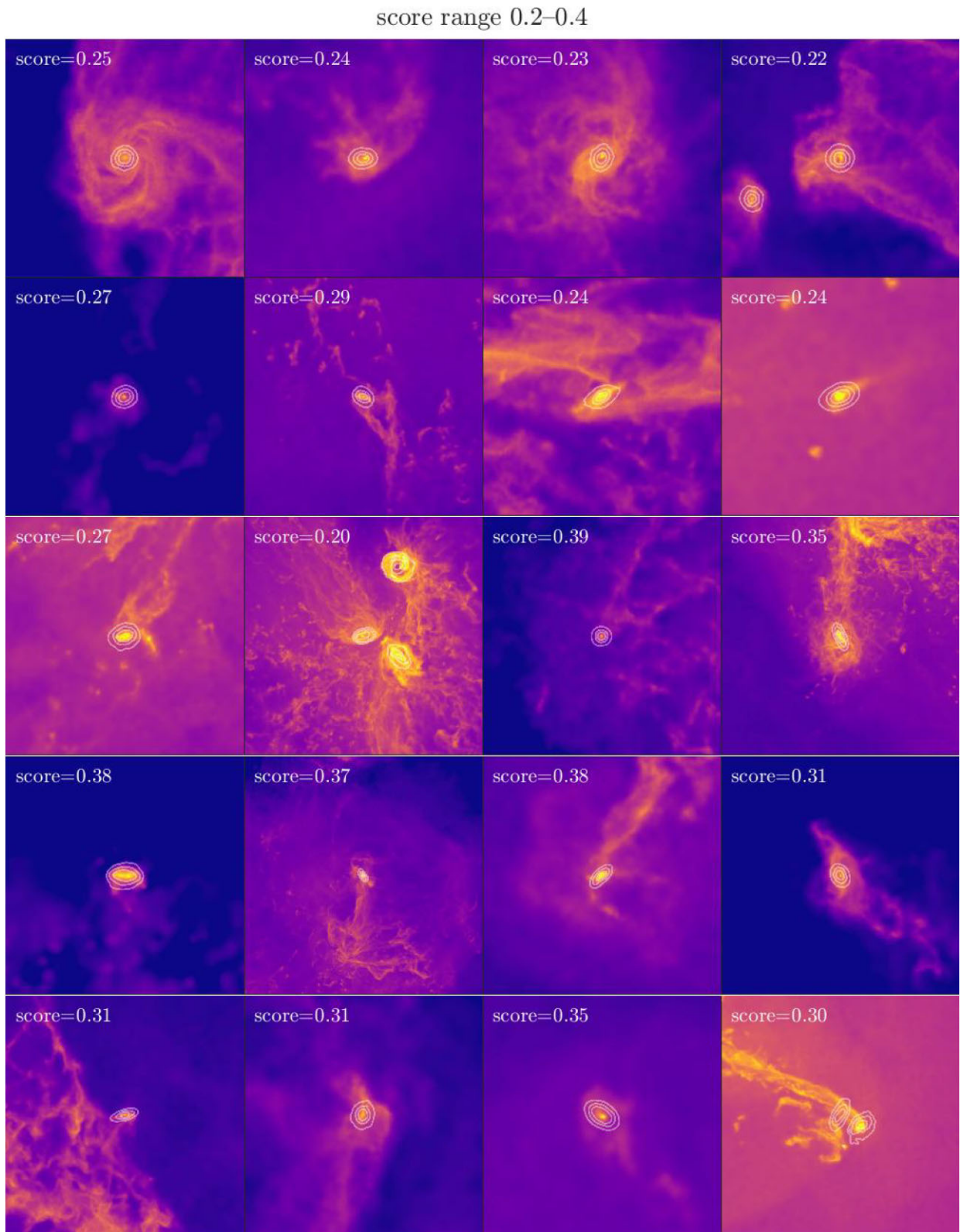


Figure B1. Sample of galaxy images of scores 0–0.2.

**Figure B2.** Sample of galaxy images of scores 0.2–0.4.

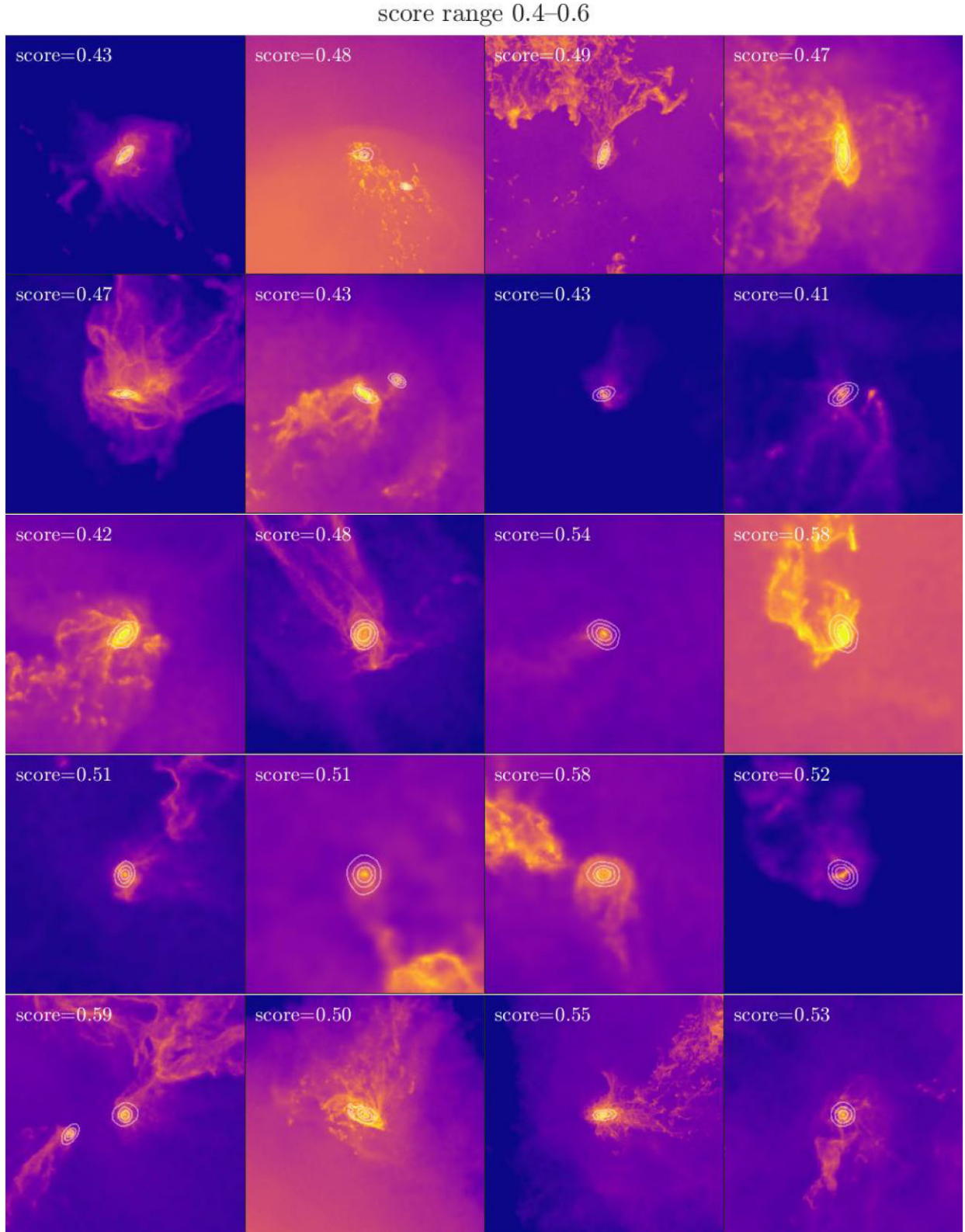
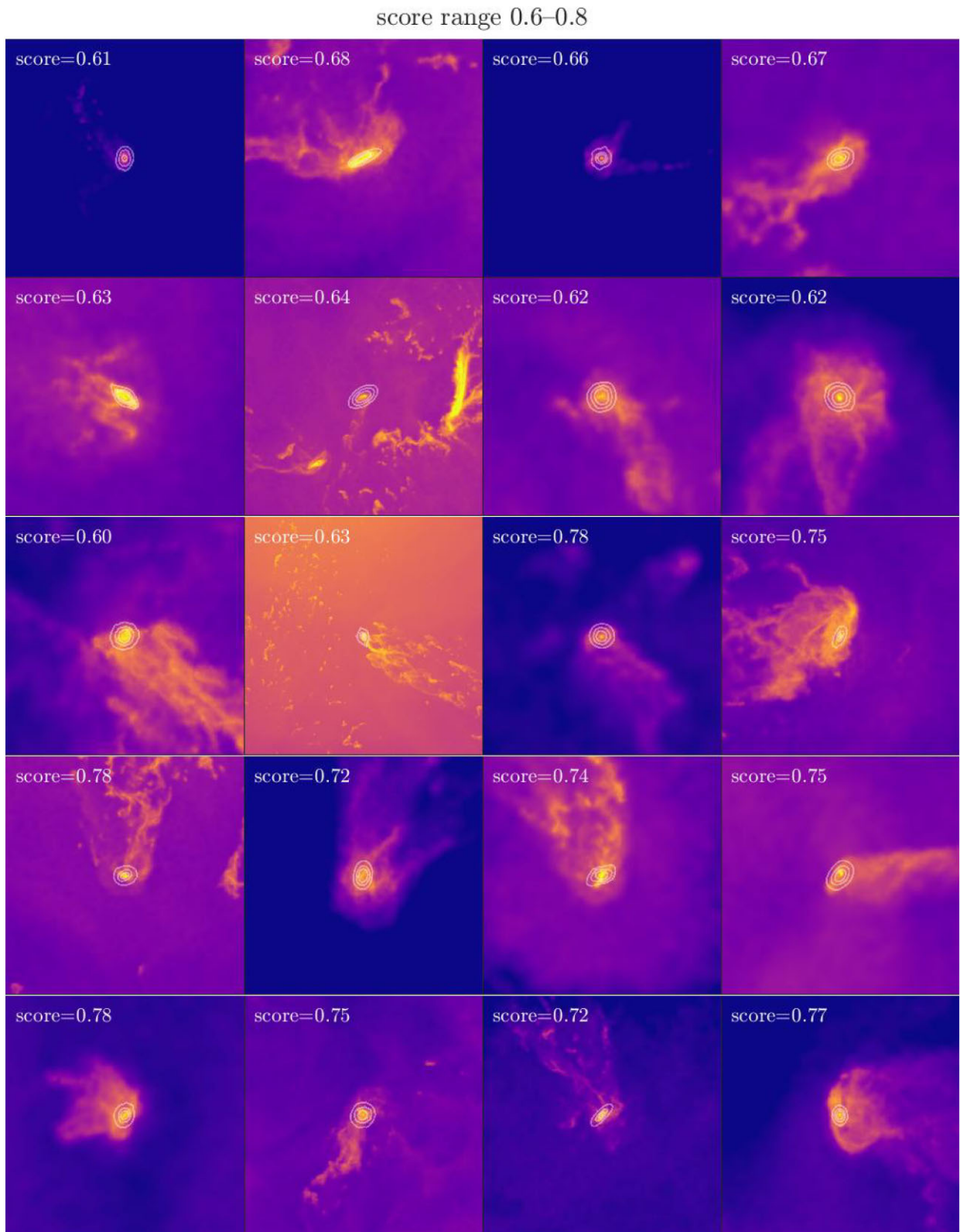


Figure B3. Sample of galaxy images of scores 0.4–0.6.

**Figure B4.** Sample of galaxy images of scores 0.6–0.8.

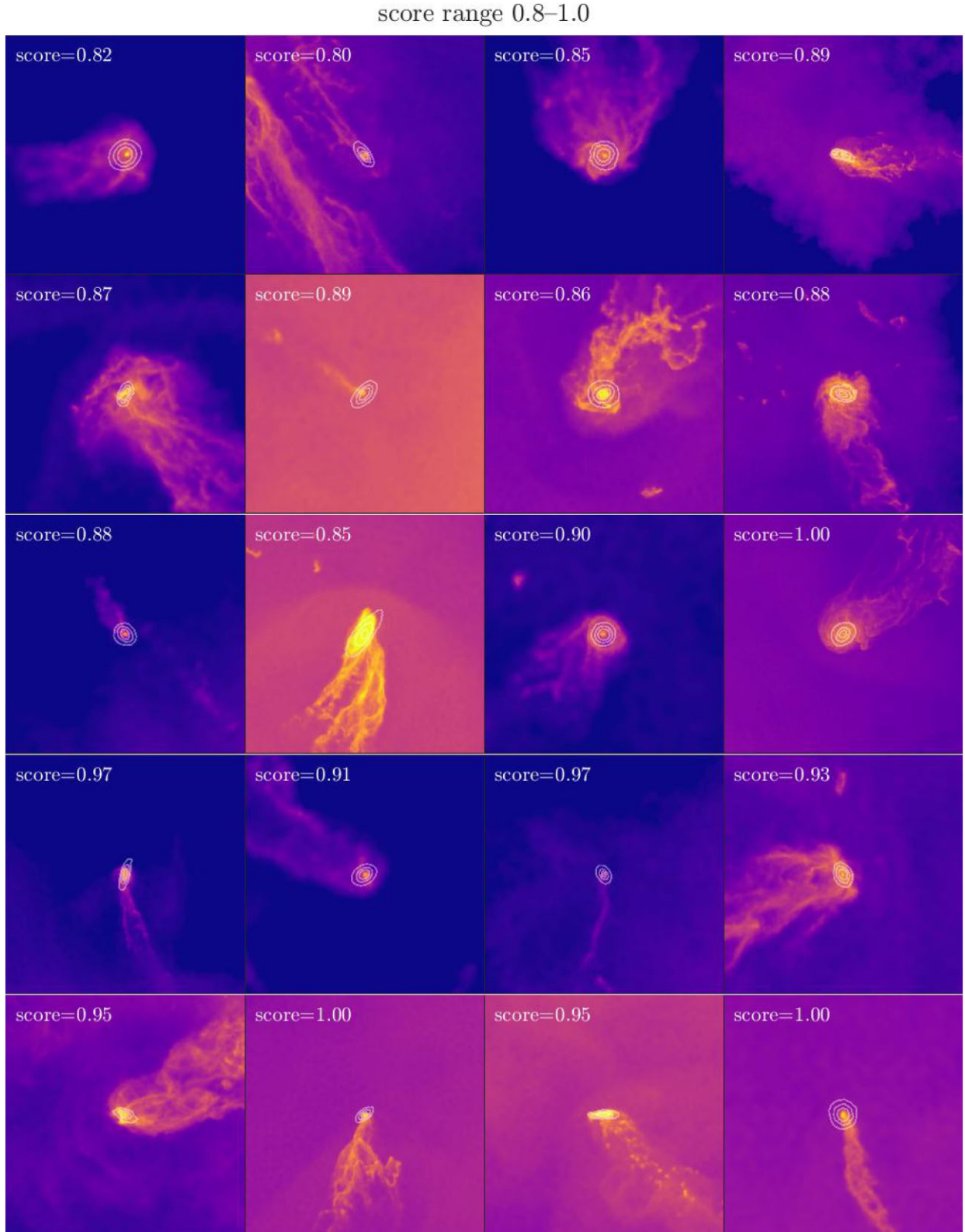


Figure B5. Sample of galaxy images of scores 0.8–1.0.

This paper has been typeset from a \LaTeX file prepared by the author.