Astronomical and Astrophysical Transactions (*AApTr*), 2023, Vol. 33, Issue 4, pp. 347–354,
ISSN 1055-6796, Photocopying permitted by license only, ©Cambridge Scientific Publishers

Galaxy origin: the first billion years of the Universe

O.K. Sil'chenko*

Sternberg Astronomical Institute of the Lomonosov Moscow State University, University av. 13, Moscow 119234, Russia

Received 6 February, 2023

Now with large telescopes we observe galaxies at redshifts of more than 5; it means that we observe them in the first billion year of the Universe expansion – very soon after the Big Bang. The galaxies were then already well-shaped, so the epoch of the galaxy formation beginning must be shifted to $z > 20$. Moreover, the quasars seen as active nuclei of distant galaxies had very massive black holes in their centers then, at $z > 6$ – mostly exceeding a billion solar masses; so the epoch of the central supermassive black hole formation must be at $z > 30$. We did not still see how the galaxies and the supermassive black holes formed; but the observational data on such early objects allow us to restrict some key physics of their formation.

Keywords: Galaxies: evolution; black holes: masses

1 Introduction

There are two experimental ways to study the evolution of galaxies. Firstly, we can observe the nearby galaxies in great detail and then try to "create" a similar model object, with all the properties known from observations, by involving various physical mechanisms into their evolution – mechanisms including dynamical effects, star formation, various element nucleosyntheses. This approach is the model-dependent one, of course: we can involve only those physical mechanisms which we know; if there are some still unknown mechanisms, the model would fail. The second approach is provided by possibilities of modern large telescopes: we can observe very distant galaxies. Due to the finite light velocity, just now we observe very distant galaxies in their state many (billion) years ago. If we assume the metric evolution of the expanding Universe, we can connect the redshift of the galaxy observed and the time which its radiation needs to reach us. The conventional LCDM-cosmology gives for the nearby Universe the age of 13.8 Gyr – just this time is needed to expand the Universe from the dimensionless point to the present size. It is for the redshift $z = 0$ – for the nearby part of the Universe. For the Universe at an arbitrary redshift, the age of expansion can be estimated through the parameters of the cosmological model as follows:

$$t(z) = \frac{1}{H_0} \int_z^\infty \frac{dz}{(1+z)E(z)} ,$$

*Email: sil@sai.msu.su

where

$$E(z) = \sqrt{\Omega_\Lambda + (1-\Omega_0)(1+z)^2 + \Omega_m(1+z)^3 + \Omega_r(1+z)^4}\,,$$

and Ω_Λ, Ω_m, and Ω_r are the mean cosmic densities of dark energy, dark matter, and relic radiation in units of critical density, $\rho_c = \dfrac{3H_0^2}{8\pi G}$. In the particular case, which, as we think, just takes place in our Universe – in the flat geometry, where $\Omega_\Lambda + \Omega_m = 1$, the integral can be solved analytically:

$$t(z) = \frac{2}{3H_0\sqrt{\Omega_\Lambda}} \ln\left(\frac{\sqrt{\Omega_\Lambda(1+z)^{-3}} + \sqrt{\Omega_\Lambda(1+z)^{-3} + \Omega_m}}{\sqrt{\Omega_m}}\right).$$

It means that within the conventional cosmological model, with $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$, we can observe almost an entire evolutionary path of galaxy population: when we observe a galaxy at the redshift $z = 0.5$, we see it five billion years ago; when we observe a galaxy at the redshift $z = 1$, we see it more than eight billion years ago... When we observe a galaxy at $z = 5$, we see it only one billion years after the Big Bang. The most distant galaxy observed now is at $z = 11$ (Oesch et al. 2016, Jiang et al. 2021), and we see it only 0.4 Gyr after the Big Bang. All the life of the Universe is directly observed. It may seem that this approach is more direct than the former proposed by us. But it is not so. At various redshifts we observe *different* galaxies, and they look differently. To connect these different galaxies with their different characteristics into a single evolutionary track, we need also models and suggestions about the physical mechanisms dominant at various stages of galaxy evolution. However, a lot of information on galaxies at various redshifts (during various Universe evolutionary stages) restricts significantly our choice of such mechanisms. However, we do not still see *the formation* of galaxies – all galaxies that we observe are already well-shaped. So as concerning the very galaxy formation, we must still believe into the theory predictions.

2 The galaxy formation and evolution within the model approaches

The development of galaxy formation theories started about 50 years ago with the models of so called "monolithic collapse". Based on the Jeans' consideration of the gravitational instability of a massive gaseous cloud, Partridge and Peebles (1967) found that the collapse of a typical galaxy-size protogalactic cloud, with the mass of 10^{11} solar masses, had to separate from the Universe expansion at the redshift of 10–30. The exact numerical simulations undertaken by Richard Larson (1969) within this framework provided a model elliptical galaxy with the valuable size and density characteristics (though rotating too rapidly) as a product of Jeans-instability collapse of a purely gaseous cloud.

However, at the end of the 1970s a cosmological idea that the Universe had to be dominated by the gravitation of *dark* matter became popular, and the key event for starting galaxy formation inside the entire expanding Universe was recognized to be Jeans-instability collapse of a *dark* halo – again, spherical in its finite shape and rotating. White and Rees (1978) proposed the following sequence of events: a dark halo collapsed under its own gravitation, baryons, completely in gaseous phase, followed the dark matter and virialized with the dark halo, filling its entire volume. The observational view of baryons following the dark matter spatial redistribution can be inspected in Fig. 1.

The equilibrium temperature of the gas virialized inside the dark haloes kept it warm–hot, and the gas radiated and cooled. Because of possessing relic angular momentum (Peebles 1969), the gas accumulates in a disk – the location of the potential minimum within the dark halo – and forms stars there. The first galaxies had to be disk galaxies, not elliptical ones. And they had to be dwarfs: the initial, at $z > 1000$, matter distribution in the

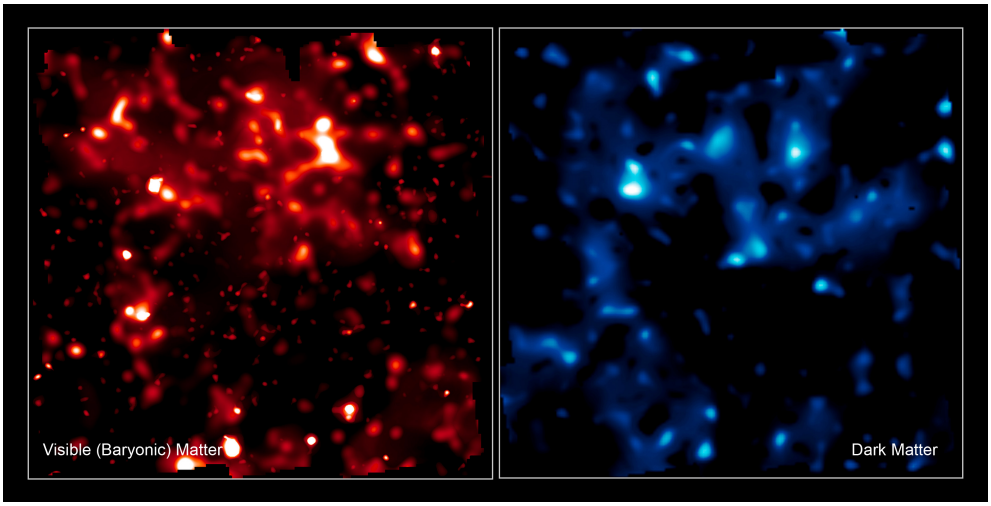


Figure 1 The observed gravitating dark matter and baryonic density in the field of the COSMOS survey (about the COSMOS project – see Koekemoer & Scoville (2005)). (Illustration courtesy of NASA/ESA Image Gallery, public collection.)

Universe was homogeneous, as we knew from the microwave background all-sky temperature distribution (e.g. Bennett et al. 2013), and to form large non-linear structures, such as Milky Way, the whole large-scale matter distribution in the Universe had to evolve long enough. Cosmologists fix the moment in their all-Universe evolution simulations when the first galaxies arise, corresponding to the epoch when the dark haloes with the masses of about 10^8 solar masses collapse (Bromm & Yoshida 2011).

After the first dwarf disk galaxy formation, the further path of galaxy evolution is represented in cosmological simulations by a sequence of violent and quiet episodes. Violent episodes are merging of two dark haloes and then merging of the galaxies inhabiting these haloes; and between two mergers the galaxy experiences quiet evolution: the hot gas distributed over the dark halo radiates, cools, accretes to the galaxy disk ("hot accretion") and feeds the consequent star formation in the disk. The major merger of two equal-mass disk galaxies (or at least of two disk galaxies with comparable masses) produces a spheroidal galaxy, or a future bulge; after the next quiet episode with stellar disk growing due to star formation fed by "hot accretion", we obtain an "early-type spiral" – a disk galaxy with a large bulge in the center.

It is a quite schematic picture. In fact, the modern scenaria of galaxy evolution includes many other mechanisms and episodes quite necessary to form a Hubble sequence of galaxy morphological types (Hubble 1936). Since "hot accretion" had to proceed all the time, the early scenaria implied that elliptical galaxies – the pure stellar spheroids – had to form through a major merger just now, a few hundred million years ago, otherwise they had time to grow a disk around their spheroids. Any major merger of two gas-rich spirals must be followed by a star formation burst, so their stellar populations must be young. Meantime, from observations we know that nearby elliptical galaxies have very old stellar populations, typically older than 10 Gyr, and the larger is the mass of the galaxy, the older is its stellar population. To explain this "downsizing", central black holes were involved. If the central supermassive black hole is fed by gas inflow, it begins to radiate, heats the gas and stops star formation. The larger the galaxy is, the larger is its central black hole, the brighter is

the radiation of the active nucleus. If we switch on active nuclei – quasars, say, at $z = 4$, we quench star formation in massive elliptical galaxies 12 Gyr ago, and then now, at $z = 0$, we obtain massive elliptical galaxies with old stellar populations.

Another important improvement was a suggestion of "cold accretion" by Dekel and Birnboim (2006). In the course of the development of the large-scale structure of the Universe, so called filaments form which contain both dark matter and baryons, and their gas is rather cold. The filaments connect the nodes of the Universe large-scale structure where galaxies inhabit. The gas of the filaments goes to the nodes where the gravitating massive dark haloes settle, and in such a way it can feed the galaxies by cold gas inflow bringing not only the gas for star formation, but also the angular (orbital) momentum to expand the galactic stellar disks. It is quite important to describe the strong growth of galactic angular momentum observed in the last 8 Gyr (Simons et al. 2017).

3 Real galaxies in the young Universe

But what do we really know about galaxies in the young Universe from observations? There are few methods to search for distant galaxies in photometric wide-field surveys, and all of them suffer strong astronomical selection biases being intended to select galaxies with a narrow range of properties.

The most popular method is Ly-break one. It was invented by Charles Steidel in the 1990s (Steidel et al. 1996). The basic physics of the method is a suggestion that a galaxy forms its stars very recently, and hence it has a rich envelope of neutral hydrogen (HI) around starforming sites. Then the integrated spectrum of its young stellar population must be affected by transition through the optically thick HI envelope: all radiation bluer than the Ly-break ($\lambda = 912$Å in the restframe) must be completely absorbed, but in the wavelength larger than $\lambda = 912$Å in the restframe the galaxy must be bright and blue due to its young stellar population. In broad-band photometry it looks like the galaxy sudden appearance in some band while it is absent in the nearby bluer filter. The "Ly-break" galaxies at the redshift of $z = 3$ are absent in the filter "U" while they are bright in the filter "B", and it has been the first discovered and studied large sample of the distant galaxies (Shapley et al. 2001). To search for more distant galaxies, we need redder filters in our photometric surveys, because the Ly-breaks are shifted redwise by the galaxy redshift. For example, galaxies at $z = 7$ are not seen in the optical spectral range at all, but appear as a bright object in the near-infrared filter "J" (Fig. 2).

Now "Ly-break" galaxies are found till $z = 11$ (the Universe age of 0.4 Gyr); at $z > 9$ (the Universe age of 0.5 Gyr) a small sample of a dozen galaxies are found, and at $z > 5$ (the Universe age of 1 Gyr) hundreds of "Ly-break" galaxies are studied (Salmon et al. 2018) allowing us to derive the statistics of their stellar population properties. The properties of "Ly-break" galaxies are very homogeneous between $z = 8$ and $z = 3$: the stellar mass of 10^{10} solar masses, the star formation rate (SFR) of a few dozen solar masses per year, and a metallicity slightly subsolar, a factor of 2 in average. Since we do not observe any evolution of Ly-break galaxies between $z = 8$ and $z = 3$, in the first two billion years of their life, despite the noticeable level of SFR, we must accept that the Ly-break method selects only galaxies with very certain properties. By applying only this method, we cannot search, for example, for massive galaxies, with matured structure. Our knowledge about the galaxy population in the first Gyr of the Universe life must be strongly incomplete.

However, we recognize some massive galaxies in the epoch close to this: at redshifts till 4.5 sub-millimeter galaxies are found in a number sufficient for statistics. The range of properties for this class of galaxies selected by searching in submillimeters is also rather narrow. We know that in the nearby Universe the galaxies with the most intense star formation have the

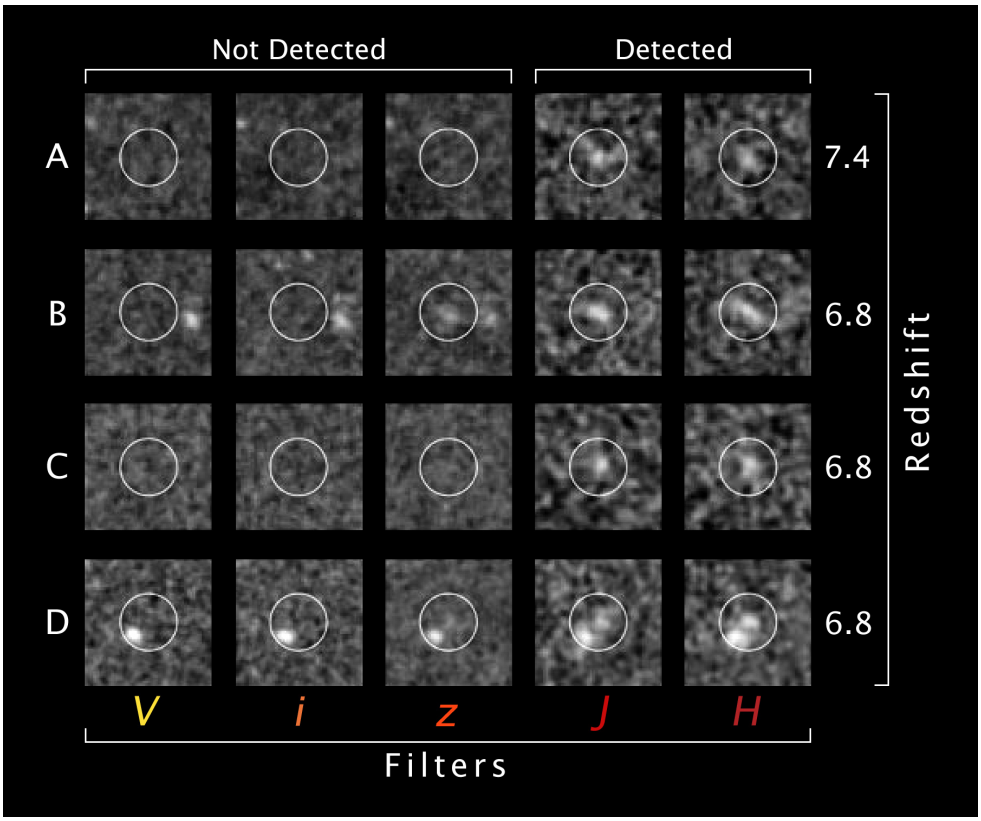


Figure 2 Ly-break galaxies at $z = 7$ through different filters. (Illustration courtesy of NASA/ESA Image Gallery, public collection.)

peak of their spectral energy distribution at $\sim 100\,\mu$m. To search for such galaxies at $z > 3$, we must observe the submillimeter spectral range, $\lambda > 400\,\mu$m. The targeted instruments were constructed, and the targeted wide-field surveys were undertaken, firstly at Hawaii, at the mountain higher than 4 km, in very dry atmosphere (Smail & Ivison 1997). A large sample of submillimeter galaxies was found, and their properties were extreme: the stellar mass of 10^{11} solar masses, the star formation rate (SFR) of a few hundred solar masses per year, up to 2000–3000$M_\odot$/yr, and all this enormous star formation is confined in the dust cocoon with the radius of less than 1 kpc. The first idea was that submillimeter galaxies were massive elliptical galaxies in the moment of their formation: the derived masses and sizes matched the characteristics of elliptical galaxies at $z = 2$ (Trujillo et al. 2007), and the intense star formation promised that it would be brief. But the discoveries of the last years put this hypothesis under doubt.

Now submillimeter galaxies are searched for mostly at the southern telescopes: in Antarctica, at the South Pole Telescope (SPT), and in the Atacama desert, in Chile, at the altitude of more than 5 km, at the Atacama Cosmology Telescope (ACT). These submillimeter telescopes compile lists of candidates to high-redshift super-starforming galaxies, and then the submillimeter-millimeter interferometer ALMA provides their redshifts and images, both in dust continuum and in gas emission lines – the [CII] line ($\lambda = 158\mu$m in the restframe) and CO lines (a set of different-excitation rotational transitions). Recent studies of the submil-

limeter galaxies with the ALMA (Rizzo et al. 2020, Rizzo et al. 2021) have revealed that the structure is rather complex: while dust continuum emitted by the dust heated with star formation burst is very compact and confined to the central region of a galaxy, $R < 1$ kpc, beyond this hot star-forming cocoon some extended cold gaseous disks are observed traced with [CII] and CO lines up to several kiloparsecs from the center. These disks demonstrate regular plane circular rotation, and their immense rotation velocities corresponding to their large stellar masses exceed the gas-cloud velocity dispersion by an order. Such rotation-supported galaxies do not relate in any sense to possible elliptical descendants and have not been expected in any cosmological simulations. If it will be found that such massive rotation-supported disk galaxies are common in the first Gyr of the Universe life, it would create a lot of problems for the theoretical views on galaxy evolution.

4 Supermassive black holes in the hearts of bulge-dominated galaxies

In the nearby Universe most galaxies with a large spheroidal component – a bulge – possess a supermassive black hole in their centers. Some (a few) of these supermassive black holes can be observed by their immense radiation exceeding the whole luminosity of their host galaxy – so called "active galactic nuclei" (AGN). But to provoke a phenomenon of AGN, one must provide a fuel for it, the gas inflow into the black hole's gravitation radius. It is a difficult task: some massive galaxies (ellipticals) do not have enough gas, some massive galaxies (spirals) have gas in their disks, but this gas has a high angular momentum and so cannot reach the very center of a galaxy. So the phenomenon of AGN is very rare, and the most supermassive black holes in the centers of galaxies are "sleeping": we can detect them only by their immense gravitation. However their gravitational effect onto gas and star dynamics in the center of a galaxy allows us to measure their masses though we do not see them directly: we can estimate the contribution of the stellar component into the whole gravitational potential by using its photometry, and then the difference between the observed $\sqrt{v_{rot}^2 + \sigma_*^2}$ and the stellar contribution would be related to gravitation (mass) of a black hole. After some statistics have been collected for the nearby galaxies, a tight proportional relation has been found between the mass of the central black hole and the mass of galactic bulge: $M_{BH} = 0.003 M_{bulge}$ (Magorrian et al. 1998, Gebhardt et al. 2000, McConnell & Ma 2013) with a fairly small scatter.

Such an impressive tight relation has stimulated numerous scenaria of galaxy evolution with so-called "co-evolution" of the bulges and central black holes (e.g. Hopkins et al. 2008, Somerville et al. 2008). Indeed, according to cosmological simulations, galaxies must merge permanently. After a major merger of two disk galaxies, we obtain a spheroidal galaxy (Barnes & Hernquist 1992), or a bulge, with the mass of this spheroid roughly equal to the sum of the merged galaxies. But the central black holes must also merge, and the mass of the final central black hole would be also the sum of two black holes which initially populated the centers of merged galaxies. In such a way, a synchronous proportional growth of the masses of bulges and black holes is provided over the whole history of the Universe. However, it must be very slow: to follow the hierarchical growth of dark haloes, we must recognize that, say, at $z = 3$ typical galaxy masses were to be of order of 10^{10} solar masses, and then typical black holes masses had to be of $3 \cdot 10^7$ solar masses; and the larger the redshift, the smaller the masses of the central black holes. But we can check it with observations! The most distant quasar is observed at $z = 7.5$, and about 200 quasars are known now at $z > 6$. And many of them have black hole masses (instead, the mass low limit due to the Eddington luminosity limit) of a billion solar masses! The age of the Universe at $z > 6$ is less than one Gyr. Within a standard cosmological scenario of synchronous growth of the central black hole and its host galaxy, the quasars have no time to grow their black holes to a billion

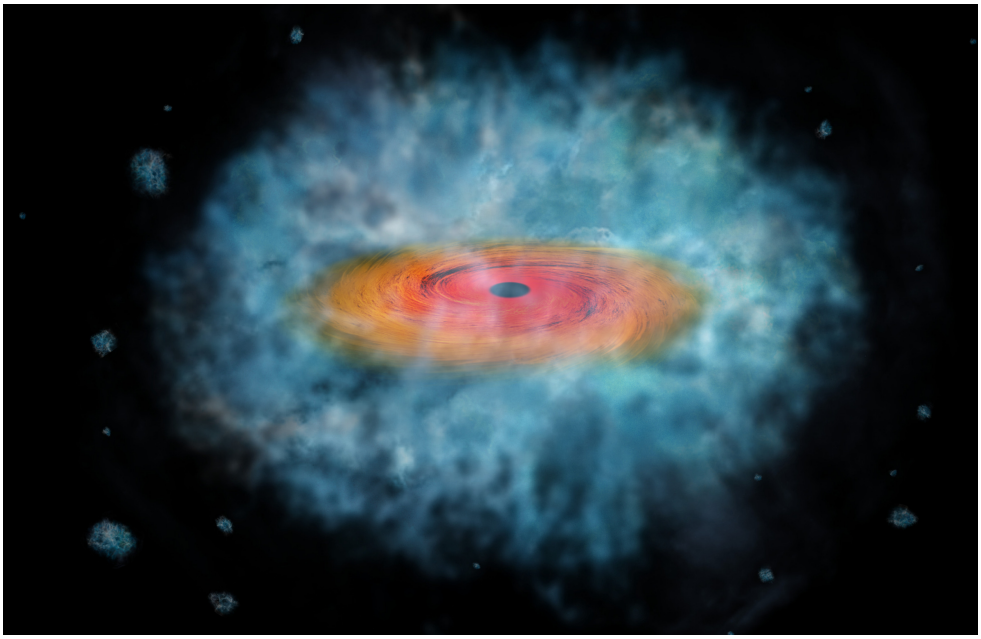


Figure 3 An artist's impression of the "seed" black hole formation inside a small dark halo filled by warm gas. (Illustration courtesy of NASA/ESA Image Gallery, public collection.)

solar masses. In addition, the interferometer ALMA was lucky to measure rotation speed and dynamical masses of the host galaxies for a few dozen of quasars at $z > 6$ by mapping the emission line [CII]$\lambda 158\mu$m; the ratio of the central black hole masses to the dynamical masses of their host galaxies at $z > 6$ has appeared to be several times larger than that for the nearby galaxies (Decarli et al. 2018, Tripodi et al. 2022). It becomes evident that within the first billion years of the Universe expansion the black holes grew much more rapidly than the galaxies. It was a challenge for the theorists, and they have responded decently.

A set of scenaria were proposed which intended to cultivate a supermassive black hole with the mass of a billion solar masses alone, without any galaxy around it, for one billion years. The main feature of all scenaria is the existence of "seed" black holes, with masses of 10^3–10^6 solar masses, at the redshift of 40–50 (Fig. 3). Then a supermassive black hole with the mass of a billion solar masses can develop till $z = 6$ by simple disk gas accretion at the Eddington limit within a small dark halo. The problem is the origin of the "seed" black holes. In principle, they can be remnants of fast evolving first-generation stars ("Population III") with a null metallicity which are expected to be more massive than the stars in the present epoch. There are physical constructions which allow us to give a birth even to a (proto-)star with the initial mass of 10^6 solar masses (Umeda et al. 2016). Moreover, it is possible to organize a "monolithic collapse" of all the gas filling the dark halo into a central black hole: the necessary condition for this is heating of the gas in the process of accretion till the temperature exceeding 10^4K – to prevent gas cloud fragmentation and consequent star formation over the extended disk. Wise et al. (2019) propose the neighboring starforming galaxies as a source of heating: so in some dark haloes with the mass of 10^8 solar masses the galaxies without black holes form, and in others – the black holes without galaxies.

5 Summary

The cosmological theory promised us that the early galaxies would be dwarf disk galaxies, with frequent mergers between them accompanied by star formation bursts and galaxy transformation into spheroids. What do we see in the real Universe? Indeed, most galaxies in the first billion years after the Big Bang are disks. However, among the galaxies with the most intense star formation at $z > 4$ we observe rather massive objects, with the stellar mass of 10^{11} solar masses which must be absent or must be extremely rare in that epoch. Moreover, their extended gaseous disks are dynamically cold demonstrating very low velocity dispersion of the gas clouds; it means that they do not experience any mergers in their short lives. The unexpected finding was also that supermassive black holes preceded the very formation of galaxies: the quasars observed at $z > 6$ have their black hole masses of more than 10^9 solar masses each of which could be formed only if no star formation proceeded around them in that early epoch.

References

Banados E et al. *Nature* **553** 473 (2018)
Barnes J E & Hernquist L *Ann. Rev. A & A* **30** 705 (1992)
Bennett C L et al. *ApJ Suppl. Ser.* **208** Aid.20 (2013)
Bromm V & Yoshida N *Ann. Rev. A & A* **49** 373 (2011)
Decarli R et al. *ApJ* **854** Aid.97 (2018)
Dekel A & Birnboim Y *MNRAS* **368** 2 (2006)
Gebhardt K et al. *ApJ* **539** L13 (2000)
Hopkins Ph F et al. *ApJ Suppl. Ser.* **175** 356 (2008)
Hubble E P *Realm of the Nebulae* (New Haven: Yale Univ. Press, 1936)
Inayoshi K, Visbal E, Haiman Z *Ann. Rev. A & A* **58** 27 (2020)
Jiang L et al. *Nature Astron.* **5** 256 (2021)
Koekemoer A M & Scoville N Z *New Astron. Rev.* **49** 461 (2005)
Larson R B *MNRAS* **145** 405 (1969)
Magorrian J et al. *AJ* **115** 2285 (1998)
McConnell N J, Ma Ch-P *ApJ* **764** Aid.184 (2013)
Oesch P A et al. *ApJ* **819** Aid.129 (2016)
Partridge R B & Peebles P J E *ApJ* **147** 868 (1967)
Peebles P J E *ApJ* **155** 393 (1969)
Rizzo F et al. *Nature* **584** 201 (2020)
Rizzo F et al. *MNRAS* **507** 3952 (2021)
Salmon B et al. *ApJ Lett.* **864** L22 (2018)
Shapley A E et al. *ApJ* **562** 95 (2001)
Simons R, Kassin S, Weiner B J, et al. *ApJ* **843** Aid.46 (2017)
Smail I, Ivison R J, Blain A W *ApJ* **490** L5 (1997)
Somerville R S et al. *MNRAS* **391** 481 (2008)
Steidel C C et al. *ApJ Lett.* **462** L17 (1996)
Tripodi R et al. *A & A* **665** A107 (2022)
Trujillo I et al. *MNRAS* **382** 109 (2007)
Umeda H et al. *ApJ* **830** L34 (2016)
White S D M, Rees M J *MNRAS* **183** 341 (1978)
Wise J H et al. *Nature* **566** 85 (2019)