The origins of massive black holes

Marta Volonteri1,2, Mélanie Habouzit2,3 and Monica Colpi4,5

Abstract | Massive black holes (MBHs) inhabit galactic centres, and power luminous quasars and active galactic nuclei, shaping their cosmic environment with the energy they produce. The origins of MBHs remain a mystery, and the recent detection by LIGO/Virgo of a black hole of almost 150 solar masses has revitalized the questions of whether there is a continuum between ‘stellar’ and ‘massive’ black holes, and what the seeds of MBHs are. Seeds could have formed in the first galaxies or could be related to the collapse of horizon-sized regions in the early Universe. Understanding the origins of MBHs straddles fundamental physics, cosmology and astrophysics, and bridges the fields of gravitational-wave physics and traditional astronomy. With several existing and upcoming facilities in the next 10–15 years, we foresee the possibility of discovering the avenues of formation of MBHs. This Review links three main topics: the channels of black hole seed formation, the journey from seeds to MBHs, and the diagnostics on the origins of MBHs.

We highlight and critically discuss current unsolved problems, touching on recent developments.

Observations reveal that astrophysical black holes come in two distinct families: stellar black holes, relics of massive stars, widespread in all galaxies of the Universe; and massive black holes (MBHs) at the centres of most of the galaxies in today’s Universe, with masses up to several billions of solar masses. MBHs fill the upper end of the mass spectrum of black holes, which at the lower-mass end is populated by stellar black holes. The lightest MBH candidate that has a dynamical mass estimate1, in the galaxy NGC 205, has a mass of 6,800 $M_\odot$, whereas the heaviest stellar black hole2, GW190521, resulting from a merger, has a mass of almost 150 $M_\odot$. MBHs are central actors in the ecology of galaxy formation through the energy produced as they grow and shine as quasars3. The presence of quasars powered by billion-solar-mass MBHs when the Universe was less than a billion years old4 requires at least some MBHs to have grown spectacularly in a short time. The ubiquity of MBHs in galaxies such as the Milky Way and the presence of low-mass MBHs in some dwarf galaxies today point to widespread MBH formation, but also suggest that some MBHs have not grown much since their birth. Explaining all these elements in a single coherent picture remains a challenge for theoretical models. MBHs must have originated from black hole seeds that have grown in mass over time (BOX 1): observations tell us that most MBHs have grown in mass by accretion by orders of magnitude5, and it is implausible to form stars with masses up to billions of solar masses leading directly to the observed MBH mass distribution. The seeds may well have been born with masses straddling6 the two families of black holes, in what is currently an observational ‘desert’ but could be filled by future observations.

The seeds can have primordial origin, generated by the collapse of high-contrast density perturbations in the early Universe, possibly as early as at the time of inflation, when the Universe experienced a brief phase of rapid exponential expansion. Inflation is also pivotal for generating cosmic structures. The small density perturbations generated by quantum fluctuations were accentuated by gravity over time: since dark matter dominates the total mass budget in the Universe, initially these overdensities generated dark matter halos within which baryons flowed. When the baryonic gas in dark matter halos became sufficiently dense and cold, star formation was initiated and galaxies were born. High initial overdensities eventually formed massive galaxies and clusters of galaxies, whereas underdensities evolved into voids. Seeds could then have been generated in galaxies, from the collapse of massive stars or from mergers of stars or stellar black holes. MBHs can reveal themselves directly only when they grow in mass: either by merging with another black hole, releasing gravitational waves (GWs); or by accreting gas or stars from their surroundings, when light is emitted as matter nears the horizon. Constraining the origins of MBHs therefore requires tracing their mass growth through the whole cosmic time theoretically and observationally in a truly multi-messenger effort (BOX 2).

Channels of black hole seed formation

The physical processes leading to the formation of MBHs from ‘gas clouds’ were first outlined by Martin Rees in 19785. Most of the formation pathways’ studied today are still based on these seminal ideas. A seed is a black hole of yet unconstrained initial mass, in the
range from a few hundreds to the order of a million solar masses. Seeds with masses between \( \sim 10^2 \, M_\odot \) and a few \( 10^4 \, M_\odot \) are commonly referred to as ‘light’ and those of \( 10^5 \, M_\odot \) to \( 10^6 \, M_\odot \) as ‘heavy’. Seeds grew into MBHs if they experienced a sufficient mass increase during the growth of dark matter halos and galaxies.

Although the presence of MBHs in today’s galaxies is ubiquitous, this may not have been the case at the time of seed formation. With time, the assembly of cosmic structures modified the occupation fraction (that is, the fraction of galaxies and halos hosting an MBH) as they evolved: faster-growing halos have a chance of hosting an inherited seed as the result of a merger. This enhances the MBH occupation fraction over time. Below, we describe the astrophysical mechanisms of seed formation, discussing pros, cons and bottlenecks, with the warning that the surveyed channels for forming seeds are not mutually exclusive.

### Channels linked to gas and stars

Except for traces of lithium, no elements heavier than hydrogen and helium were synthesized during Big Bang nucleosynthesis, so that the first stellar objects originated within gas clouds of primordial composition cooling inside the lowest-mass, earliest-forming dark matter halos. Further-generation stars formed in heavier halos, inside gas clouds enriched by traces of heavy elements released in supernova explosions. The lack of elements heavier than helium, called metals, provides favourable conditions for the birth of seeds. Because of fewer avenues for energy exchanges and spectral transitions, cooling is less efficient, and gas has higher equilibrium temperatures. The concept of Jeans mass sets the scale of collapsing gas clumps that give rise to stars: gravity overwhelms gas pressure in clumps with mass above the Jeans mass, which increases with temperature. Although this classic argument should be taken with caution\(^\text{18}\), in the absence of metals, hotter Jeans-unstable clouds fragment less so that the stars that form are more massive. The same argument about spectral lines means less effective radiation pressure, and at metallicities lower than 0.01 the solar value, suppression of winds that carry away mass from stars leads to relic black holes with mass similar to that of the progenitor star\(^\text{19,20}\). Low metallicity is therefore, in principle, doubly favourable as it leads to heavier stars and heavier black holes at fixed stellar mass.

### Single very massive or supermassive stars

The most common theories of seed formation are thus based on the collapse of stellar-like objects formed in metal-free and metal-poor environments\(^\text{21}\). The first stars, dubbed Population III stars, originate in metal-free clouds hosted in halos with mass \( \sim 10^5 \, M_\odot \) where the only coolant is hydrogen in its molecular state. Initially, Population III stars were thought to live short, solitary lives with one massive star (more than \( 10^5 \, M_\odot \)) forming in each dark matter halo\(^\text{22}\). Later theoretical studies found instead that several stars form per halo, with a broad mass spectrum from \( \sim 10 \, M_\odot \) up to \( \sim 10^3 \, M_\odot \) (refs\(^\text{14,16}\)), making their role as seeds more questionable, although one expects to find Population III relics in almost all galaxies\(^\text{23}\). For instance, in a galaxy like the Milky Way, models of Population III formation\(^\text{24}\) predict about \( 3 \times 10^4 \) relics, but with an average mass of about \( 10^2 \, M_\odot \), although this depends on the chosen mass distribution at birth of the stellar population.

An alternative pathway is the so-called direct collapse scenario where most of the gas in a halo of mass \( \sim 10^5 \, M_\odot \) contracts coherently and forms a single supermassive star (SMS) of \( 10^6-10^9 \, M_\odot \) that then collapses to a massive seed\(^\text{25,26}\), possibly via an intermediate phase as a quasi-star where the power comes from accretion onto a growing small black hole\(^\text{27}\) rather than nuclear burning. The key to reach such high masses is rapid growth of the protostar on timescales shorter than its thermal (Kerwin–Helmholtz) timescale, guaranteeing that it remains cold at its surface\(^\text{28}\). As a consequence of the cold surface, ultraviolet radiation that can ionize, heat and rarely gas is absent, so that accretion can continue. These conditions are favoured in halos whose gas is metal-free and able to form atomic hydrogen, but no molecular hydrogen, which triggers collapse of most of the gas in a short time. Since molecular hydrogen usually forms first, the presence of external photodissociating radiation\(^\text{29,30}\) or mechanical energy sourced by the rapid mass growth of the halo\(^\text{31}\) is needed, thus requiring somewhat contrived circumstances. This makes the sites favourable for such a process relatively rare\(^\text{32}\), with the predicted number density in the range \( 10^{-9}-10^{-7} \, \text{cm}^{-3} \), depending on the assumptions made. To invoke direct collapse to explain \( z > 6 \) quasars (\( z \) is the redshift), the average density of seeds is not a sufficient criterion: the loci favourable for SMS formation are often in satellites of massive halos, and only the few with short sinking time to the centre of the main halo are viable\(^\text{33}\). Recent simulations have also shown that SMSs may not be as massive as initially predicted but reach only a few thousands of solar masses\(^\text{11}\), because the environment is turbulent, and accretion on the protostar does not last long enough to create an SMS. Furthermore, the role of angular momentum on forming a single SMS versus fragmenting into multiple, less massive stars has not been thoroughly explored yet.

Alternatively, very rare major mergers of metal-enriched, very massive galaxies at redshift \( z \sim 8 \) have
also been advocated to be sites of SMS formation. Such collisions temporarily enhance the rate of central gas inflow by removing angular momentum via strong tidal torques, possibly preventing cooling in the centremost regions34, although no ab-initio cosmological simulation including all required physics has shown the occurrence of this process.

In summary, low-mass seeds are common, but high-mass seeds are rare and depend on processes that have not been observed yet.

Runaway stellar mergers in young, dense clusters. This channel forecasts the birth of seeds from the collapse of very massive stars (VMSs) of ~200–10^4 M⊙ forming in dense, metal-poor stellar clusters undergoing core collapse on timescales shorter than the life of the most massive stars31. The dramatic, albeit temporary, growth of the central density (up to values of 10^5–7 M⊙ pc^−3) sparks violent few-body interactions among single massive stars (or even protostars) and stellar binaries, ending with the formation of a VMS through runaway collisions35–37. The large cross-section that binaries offer and the excitation of large eccentricities during the stellar encounters lead to the physical collision of stars that repeats in a stochastic fashion until the encounter time becomes longer than a few Myr and the VMS collapses into a black hole of comparable mass38,39. VMSs can further form in gas clouds mildly polluted with metals by competitive accretion onto few central massive stars or stellar mergers37–39.

Dense, young clusters as well as more massive, compact nuclear star clusters in high-redshift galaxies appear to be ideal grounds to nurture VMSs40–41, although the effect of mass loss at collision in determining the true mass of a VMS is not fully quantified. Seeds with birth masses around 10^2–10^3 M⊙ may start forming ~300 Myr after the Big Bang40. Collisions between dark matter halos of a few 10^3 M⊙ can also provide the conditions needed to form massive clusters with a VMS41. Overall, the requirement to have mild metal pollution to spark fragmentation inside ultra-compact clusters where VMSs form leaves a wider time window for seed formation to happen, as long as the metallicity remains sufficiently low that stellar winds are weak and the collapse of the star results in a black hole of similar mass. Metallicity becomes unimportant if a stellar black hole grows into a seed through runaway tidal capture of stars42, extending seed formation to lower redshifts.

Hierarchical black hole mergers. Black hole relics of massive stars in stellar clusters of 10^5 up to 10^7 M⊙ may undergo repeated mergers to grow to the intermediate scale43. Owing to their larger mass compared with the mean mass of stars, they sink by dynamical friction at the centre of the cluster where, owing to the higher densities, they pair via binary–single dynamical encounters, building a nested core of binary and single black holes. Then, through repeated flybys and exchanges, the binaries reduce their separation and may reach coalescence. However, as gravitational binding energy is traded for kinetic energy in a dynamical encounter, such interactions can eject binaries before they merge, thus aborting the process of growth44. Their retention and merger rate depend on their mass at birth with the heavier, with masses in excess of 50 M⊙, being the most favoured for the growth via repeated mergers with other black holes45. A remaining obstacle to grow a seed by multiple generation mergers is the GW-induced recoil that the coalescence product receives46. Despite this, many studies47,48 show that the hierarchical growth is possible in nuclear star clusters with escape velocities in excess of 10^5 km s^−1.

Major gas inflows in a metal-poor, dense star cluster can further help in triggering a chain of multiple mergers, as inflows deepen the gravitational potential well, leading to the formation of seeds with masses up to 10^3 M⊙49,50. In dense, gas-rich nuclear star clusters, a single stellar black hole can also grow to become a seed if in its motion it captures low-angular-momentum gas so that quasi-spherical accretion can proceed unimpeded48.
The quest to constrain the origins of massive black holes (MBHs) has already started, and it will intensify in the future with the combination of several instruments, techniques and diagnostics. In the figure, key events in the evolution of the Universe (right) are connected to the facilities with which we can observe MBHs around some of those times, with the redshift shown on the left corresponding to the time shown on the right. The MBH mass–redshift plane (left) is overlaid with the reach of various instruments: the signal-to-noise curves for gravitational-wave (GW) facilities (for non-spinning binaries with mass ratio 0.5) as well as the approximate ranges for various telescopes and satellites. The GW antenna LISA in space and the 3G interferometers on Earth, such as the Einstein Telescope (ET)\textsuperscript{149}, will explore the mergers of black holes with masses of $10^5–10^9 \, M_\odot$. These instruments will survey the entire sky, in area and depth, from the surroundings of our Milky Way to the first hundreds of Myr of the Universe. Complementary electromagnetic (EM) missions will probe the site of accreting MBHs when seeds were still young (400–900 Myr). In the optical/near-infrared, Nancy Grace Roman Space Telescope (Roman) and Euclid will map the most luminous MBHs up to redshift $z \sim 10$. The James Webb Space Telescope (JWST) will characterize high-redshift MBHs up to $z \sim 10$, from luminous quasars to potentially newly formed MBHs. Next-generation X-ray missions (Athena, LynX, AXIS) aim at finding MBHs close to the redshift of their formation. Extremely large telescopes (such as the Extremely Large Telescope (ELT)) on Earth, with mirrors up to ~40 m diameter, will enable MBH masses to be measured in a variety of galaxies, including dwarfs, which are expected to have recorded key aspects of MBH formation. Tidal disruption events from transient surveys, such as Legacy Survey of Space and Time (LSST) with the Vera C. Rubin Observatory (Rubin), and extreme mass ratio inspirals from LISA will also extend the demography of low-mass MBHs. Multi-band and multi-messenger (GW + EM) observations will help maximize the scientific return.

The left panel is adapted from Valiante, R. et al. Unveiling early black hole growth with multifrequency gravitational wave observations. Mon. Not. R. Astron. Soc. 500, 4095–4109 (2021), REF\textsuperscript{151}, by permission of Oxford University Press on behalf of the Royal Astronomical Society. The right panel is adapted with permission, courtesy of NASA/WMAP Science Team.

**Box 2 | Diagnostics for formation/evolution of massive black holes**

The quest to constrain the origins of massive black holes (MBHs) has already started, and it will intensify in the future with the combination of several instruments, techniques and diagnostics. In the figure, key events in the evolution of the Universe (right) are connected to the facilities with which we can observe MBHs around some of those times, with the redshift shown on the left corresponding to the time shown on the right. The MBH mass–redshift plane (left) is overlaid with the reach of various instruments: the signal-to-noise curves for gravitational-wave (GW) facilities (for non-spinning binaries with mass ratio 0.5) as well as the approximate ranges for various telescopes and satellites. The GW antenna LISA in space and the 3G interferometers on Earth, such as the Einstein Telescope (ET)\textsuperscript{149}, will explore the mergers of black holes with masses of $10^5–10^9 \, M_\odot$. These instruments will survey the entire sky, in area and depth, from the surroundings of our Milky Way to the first hundreds of Myr of the Universe. Complementary electromagnetic (EM) missions will probe the site of accreting MBHs when seeds were still young (400–900 Myr). In the optical/near-infrared, Nancy Grace Roman Space Telescope (Roman) and Euclid will map the most luminous MBHs up to redshift $z \sim 10$. The James Webb Space Telescope (JWST) will characterize high-redshift MBHs up to $z \sim 10$, from luminous quasars to potentially newly formed MBHs. Next-generation X-ray missions (Athena, LynX, AXIS) aim at finding MBHs close to the redshift of their formation. Extremely large telescopes (such as the Extremely Large Telescope (ELT)) on Earth, with mirrors up to ~40 m diameter, will enable MBH masses to be measured in a variety of galaxies, including dwarfs, which are expected to have recorded key aspects of MBH formation. Tidal disruption events from transient surveys, such as Legacy Survey of Space and Time (LSST) with the Vera C. Rubin Observatory (Rubin), and extreme mass ratio inspirals from LISA will also extend the demography of low-mass MBHs. Multi-band and multi-messenger (GW + EM) observations will help maximize the scientific return.

The left panel is adapted from Valiante, R. et al. Unveiling early black hole growth with multifrequency gravitational wave observations. Mon. Not. R. Astron. Soc. 500, 4095–4109 (2021), REF\textsuperscript{151}, by permission of Oxford University Press on behalf of the Royal Astronomical Society. The right panel is adapted with permission, courtesy of NASA/WMAP Science Team.

These formation channels are only now receiving attention, following the discovery of GW190521 (REF\textsuperscript{156}). The appealing side of this process is its milder dependence on metallicity, and therefore the possible occurrence over longer cosmic times and not only in the first galaxies. The main uncertainty is in quantifying the importance of ‘kicks’ by GWs in terminating the sequence of hierarchical growth, and this depends on the depth of the cluster potential well and on the mass and spin distributions of the black holes in the binaries. These distributions are starting to be determined, but only for the low-redshift GW sources.

**Early Universe cosmology channels**

These are formation scenarios linked to the collapse of regions in the early Universe, before galaxy formation. In this case, galaxies form around black holes, and not black holes ‘in’ galaxies: the presence of a black hole in an otherwise homogeneous Universe with small density fluctuations represents an overdensity that would have attracted dark matter (if black holes do not make up the whole of dark matter; see BOX 3) and gas, thus creating sites for further galaxy formation. The black hole would then be located at the very centre of the forming halo/galaxy and would naturally grow from accretion of the infalling material and mergers with other black holes, if they are clustered at formation. The dark matter cocoon enveloping the central black hole and attracting surrounding baryons in the deepening potential well would allow the seed to grow until energy injections from star formation and from supernova explosions, or from accretion on the seed itself, stir the gas and modulate further growth. Despite these appealing features, there is no observational evidence of these processes,
and they often require ad-hoc assumptions, for which one has relative freedom precisely because of the lack of observational constraints.

**Primordial black holes.** Primordial black holes can form in the presence of high-contrast density fluctuations, which collapse on to themselves. They are not a generic prediction of inflationary cosmology, but can be generated under specific conditions, for instance in an inflationary scenario with a peak or a very high-density tail in the power spectrum of curvature fluctuations, or if there is a sudden change in the plasma pressure, for example during reheating or phase transitions. The initial mass is generally that of the horizon at the time of collapse, and can therefore span a wide range, from well below solar mass to millions of solar masses depending on the specific mechanism triggering the collapse, for instance, from ~1 g just after inflation to around a solar mass at the quantum-chromodynamic phase transition to ~10^3 M_☉ at the time of e^+e^- annihilation. Primordial black holes are a viable candidate as MBH seeds if they have an overall low mass density, compared with the total matter density, in order not to violate existing constraints. Given that MBHs represent about 10^{-5}–10^{-6} of the total matter density, and that this number has increased by several orders of magnitude as seeds became MBHs, mass windows exist for primordial black holes to be seeds.

**Cosmic string loops.** Cosmic strings are topological defects that can form during phase transitions in the early Universe both in quantum field theories and in string theory. When cosmic strings intersect and form loops, they can collapse into black holes if the loops have sufficiently low angular momentum to contract within their masses’ Schwarzschild radius or if they accrete enough dark matter to trigger collapse. GWs from oscillating string loops impose the strictest limits on the power spectrum generated by cosmic strings. Taking into account these bounds and further assuming that the size of loops depends uniquely on the time of formation of the loop and that one loop should seed one MBH per galaxy, the resulting mass and redshift windows of interest to explain MBHs can be calculated, resulting in seed masses up to 10^5 M_☉ at z ~ 30.

**From seeds to MBHs**

The middleweight black holes whose formation channels we have sketched above are seeds for MBHs with masses millions to billions of solar masses if they experienced sufficient mass increase during the growth of galaxies and dark matter halos (FIG. 1). The growth of those seeds, ancestors of the MBHs powering quasars at redshift z ~ 6–7, is rapid, but it can be gradual for most of the population, and some of the black holes formed at early cosmic times may not germinate MBHs if conditions are unfavourable.

Seeds of 10^7–10^8 M_☉ have to fight harder to develop into fully fledged MBHs, as their gravitational sphere of influence is small compared with galactic dimensions, and it is difficult to convey low-angular-momentum mass into the region where it can be captured by the seed. If they form in low-density environments and/or are too light to sink to the centre of a galaxy potential well, their growth is immediately stunted. These conditions would not be suitable for efficient growth by gas accretion, nor by mergers with other seeds, and could lead to a considerable population of low-mass wandering MBHs in galaxies. Furthermore, Population III relics have an additional complication, as their progenitor stars produce copious ultraviolet radiation that sterilizes the environment, and no immediate accretion is possible.

Although Population III relics may have difficulty in growing large by accretion despite their ubiquity in halos, the contingent conditions leading to the formation of an SMS are more exotic and such that the rate of direct collapse could be too low to explain the presence of MBHs in most galaxies. The dynamical channels based on runaway mergers of stars or black holes are elegant and based on known physics, but few investigations have explored their role in building MBH populations, and primordial black holes and cosmic string loops are basically uncharted territory in MBH studies.

**Growth by accretion**

Most of the MBH growth in the Universe is sourced by accretion in the centres of galaxies, where gas, having net rotation, approaches the last stable circular orbit, gently spiralling down and forming an accretion disk. Light is produced around the MBH, emerging from the accretion disk, the hot corona of tenuous hot plasma surrounding the disk and sometimes from a relativistic jet produced in the interaction of the magnetized plasma with the spinning MBH. A small fraction of MBH growth can also be the product of accretion of gas stripped from stars passing very close to the MBH, thus creating transient sources called tidal disruption events. Accretion rate and conversion efficiency of gravitational
Fig. 1 | Growth of massive black holes (MBH) in galaxies. MBH mass (\(M_{\text{BH}}\)) versus the stellar mass of the host galaxy (\(M_{\text{galaxy}}\)). The blue cloud represents observational data\(^{13-15}\), whereas the curves are from simulations. Vertical lines on the right side indicate the current observational range of MBH masses in the local Universe and at high redshift \(z\). Seeds must overcome hurdles to grow into MBHs, but when fully grown they regulate further growth through the energy and momentum they inject when feeding.

Eddington luminosity

Maximal luminosity above which radiation pressure on electrons overcomes gravity on the infalling matter, under the assumption of spherical symmetry.

Active galactic nuclei (AGN)

AGN and quasars are sources powered by an accreting massive black hole. Quasars are the most luminous among AGN.

Feedback

Physical processes in which the energy/momentum output of a system (or a fraction of the output) returns to or impacts the system’s input.

Radiative efficiency \(\epsilon\)

\(\epsilon\) is the efficiency at which gravitational energy is converted into radiation. It establishes the link between the accretion luminosity \(L\) and mass accretion rate \(M\): \(L = \epsilon M \dot{C}c^2\). In geometrically thin, optically thick accretion disks around black holes, \(\epsilon \approx 0.06-0.52\), depending on the spin, with \(\epsilon \approx 0\) used as reference value. \(\epsilon\) can be lower depending on the geometry of the flow. \((1 - \epsilon)M\) gives the mass growth rate of an MBH.

Energy into radiation determine the luminosity and the rate at which MBHs grow.

An important concept is that of Eddington luminosity, which sets a maximal accretion rate, and a timescale for the e-fold increase of the MBH mass of \(\sim 45\) Myr. This concept is often used as a critical value for the maximal accretion rate onto an MBH: in reality, accretion in disks can proceed above this critical value\(^{4}\), as for instance has been observed in tidal disruption events, with the geometry of the disk changing, and photons being trapped by the high densities, unable to diffuse out of the disk.

MBHs that are not (or weakly) accreting are called quiescent, with Sgr A* in the Milky Way as a paramount example. Although they emit little light, quiescent MBHs can be studied through the orbital motion of stars and gas inside the region where dynamics and kinematics are controlled by the Keplerian potential of the MBH.

Feedback and dynamics growth hurdles.

Many physical processes can slow down the growth of seeds by diminishing the MBH gas reservoir from which they accrete (Fig. 1). These processes are related either to the physics of the MBHs themselves or to the physics of their host galaxies. Active galactic nuclei (AGN) are able to release a large amount of energy on scales ranging from sub-pc to tens of kpc, potentially impacting their whole galaxies and beyond; the most spectacular example are jets that can extend over physical scales of Mpc. The released energy or momentum couples to the gas in the vicinity of the MBH. This heats and/or depletes the gas reservoir\(^{5}\), even preventing further gas inflow\(^{6}\), and can take several forms: thermal, mechanical through relativistic highly collimated jets or non-relativistic non-collimated outflows, and electromagnetic (EM) radiation\(^{7}\). This is called AGN feedback, and it is termed ‘negative’ since it suppresses the source of energy, and can take place from birth in the case of high-accretion-rate seeds. An MBH would therefore cut off its own gas supply and stop accretion until the effect of feedback has vanished, and then a new cycle of accretion–feedback–starvation resumes.

Explosions of supernovae near the MBH’s gas reservoir can also suppress MBH accretion for similar physical reasons\(^{8-10}\). Supernova-driven winds can reach velocities of more than 200 km s\(^{-1}\) in the interstellar medium, enough to remove dense gas from low-mass galaxies with shallow gravitational potential wells. As a result, MBH growth could be delayed until their host galaxies are massive enough to retain the gas in the galaxy nuclear regions. This supernova feedback is thought to be the dominant hurdle in low-mass galaxies.

Finally, MBHs accrete more efficiently when located in the centre of galaxies, embedded in a dense gas reservoir, and have a hard time when displaced. This would be the case for MBHs too light to reach the bottom of the potential well of their host galaxies\(^{11-13}\). These MBHs could wander in their galaxies for potentially a long time, and off-centre MBHs have been detected in the local Universe\(^{14-16}\). MBHs can also be kicked out of the centre of galaxies after the coalescence with another MBH. The asymmetry of the GW emission from two merging MBHs, which depends on the mass ratio of the binary and on the spin magnitudes and relative directions, can trigger the recoil of MBHs off the remnant galaxy centre, thus delaying its growth, or their expulsion from the galaxy altogether, thus requiring the acquisition of a new seed. Triple MBH interactions can also displace them from galaxy centres.

Growth timescales for \(z > 6\) quasars.

An exceptional class of MBHs is those powering luminous quasars at \(z > 6\), currently up to \(z = 7.6\)\(^{17}\), when the Universe was less than a billion years old. These are among the most massive MBHs discovered so far: they have masses of \(10^8 - 10^{9} M_\odot\), comparable to those of the most massive quiescent MBHs in the local galaxies (Fig. 1). Although rare, about 1 per Gpc\(^3\), these objects pose a timing challenge, because a seed of mass, for example, \(10^6 M_\odot\) born at, say, \(z = 11\) or 0.42 Gyr after the Big Bang, must have grown continuously, accreting at the Eddington luminosity with high radiative efficiency, to reach \(10^8 M_\odot\) in the time elapsed between \(z = 11\) and 6.

This is a feat in light of the growth hurdles for MBHs in low-mass galaxies, but if the galaxies hosting these distant quasars represent the descendants of remarkably luminous and massive \(z = 11\) galaxies such as GN-z11\(^{18}\), then their growth would be ‘normal’ in ‘exceptional’ galaxies that became massive enough to foster MBH growth very early on. In principle, such massive galaxies as GN-z11 anchor biased regions, and therefore one would expect an overdensity of galaxies around the quasars. Observationally, there is no consensus on this\(^{19-21}\): several observed quasar environments on kpc to Mpc scales show no galaxy enhancement, and some could even be underdense. However, if one...
were to confirm that quasar fields are not statistically overdense, that would not only indicate a large variance in quasar environments6, but also possibly the power of quasars to influence star formation in their own galaxy and even in nearby galaxies through their energy and momentum input6.

Solutions to this timing problem include more massive seeds, provided that they are not too rare, a substantial contribution of MBH mergers to the mass budget or some phases of accretion at super-Eddington levels. MBH mergers, however, can have both a positive and negative impact on the mass growth because they can kick MBHs out of galaxies6,87. The possibility of super-Eddington accretion is enticing, as it is a natural expectation of accretion disk physics86 and it has been predicted to occur in high-redshift galaxies under different conditions55,88,89, including large reservoirs of gas and efficient transport of angular momentum from the galactic scale down to the black hole gravitational sphere of influence owing to turbulence89 or angular momentum cancellation from multiple accreting streams81,82, both more common at high redshift. The main uncertainty is how long super-Eddington accretion can be sustained, since it is predicted to be accompanied by powerful jets that can destroy the flow of gas from the galaxy towards the MBH on small scales83 unless the jet can pierce through the whole gas distribution without affecting it, for instance if gas is confined in an equatorial disk down to the MBH accretion flow scale84. It is unclear whether such axisymmetric conditions can be achieved, taking into account the expected turbulent nature of high-redshift galaxies.

**Growth by mergers**

MBH mergers play a secondary role in the mass budget of MBHs, but they become important when gas is not readily available60,61 or in galaxies with a rich merger history, and these conditions are generally met in low-redshift massive galaxies. The role of mergers in growing seeds at high redshift is still controversial: on the one hand, the merger rate of halos was higher at early times and the dynamical timescales were shorter because the Universe was denser; on the other hand, the lightness of the seeds and the tumultuous evolution of high-redshift galaxies makes it harder for seeds to find a companion and merge.

General relativity predicts that accelerated masses in binaries emit GWs and coalescing stellar and MBH binaries are among the loudest sources of high- and low-frequency GWs, respectively97. Below we briefly outline pathways for the formation of binaries conducive to mergers.

**In situ formation of black hole binaries.** Simulations of pristine star-forming clouds show that massive Population III stars form in binaries45,49, but there is no proof that these binaries merge in situ shortly after their formation. These mergers, when delayed in time, could contribute at the per cent level to the population detectable with advanced LIGO/Virgo100,101.

As far as VMSs are concerned, numerical simulations of young star clusters suggest that VMSs acquire stellar or black hole companions, as a natural outcome of the runaway process, and that a fraction of these binaries have coalescence times shorter than the Hubble time102. Within the direct collapse scenario, general-relativistic numerical simulations show that differentially, rapidly rotating SMSs can split, during their collapse, into a black hole binary which merges shortly after if a bar-mode density perturbation is implanted in the SMS102. This may require fine-tuned conditions, since building an SMS by accretion requires low Keplerian velocities at the surface and thus slow rotation103,104. It has been further speculated that if mild fragmentation occurs around the SMS, a companion star can form at close separations and create a binary104.

**MBH binaries in galactic collisions.** Close pairs of MBHs are a natural outcome of galaxy collisions105,106 and can be discovered when the two MBHs are accreting. Dual AGN have been observed in interacting galaxies, whereas MBHs bound in a Keplerian binary on sub-parsec scales are challenging to observe, but the number of candidate systems is increasing106,109.

To become sources of GWs, MBHs have to make a long journey before reaching coalescence, which occurs on a scale of a few micro-parsecs. In galaxy mergers of comparable mass, dynamical friction against stars leads to the formation of a bound system, a binary still surrounded by stars and gas. The final crossing into the GW-driven regime then hinges on energy transfer through scatterings of single stars impinging on the binary106, and in the presence of a gaseous circum-binary disk through the interplay between gravitational and viscous torques110. In unequal-mass galaxy mergers111,112, the tidally disrupted lighter galaxy may leave a wandering MBH, or the binary evolution may stall until a third MBH comes into play to trigger a complex three-body dynamics which may end in the merger of the two closest MBHs113. The time elapsing between galaxy collisions and MBH mergers can be lengthened in massive nuclear disks (which are the outcome of galaxy mergers) by the presence of giant gas clouds and of supernovae explosions114 that rarely the medium, and in some dwarf galaxies by their loose density structure114.

At high redshift, the dynamics of seeds inside forming galaxies is key for determining their growth via accretion and mergers: seeds need to remain anchored to the minimum of the gravitational potential well where densities are the highest. But galaxies at these redshift are turbulent, clumpy and starbursty, and their morphology continues to change: thus they lack well-defined stable centres. Under these circumstances, seeds have complex, erratic dynamics75,115.

**MBHs and galaxies**

MBHs are generally found embedded in the centre of most galaxies at the present time. In the local Universe, low-mass MBHs have been detected in dwarf galaxies1, and the heaviest MBHs are found in massive elliptical galaxies (Fig. 1). Some examples are the $5 \times 10^4 M_\odot$ MBH in the dwarf galaxy RGG 118 (REF.116), Sagittarius A* in the Milky Way with a mass120,121 of $3.6 \times 10^6 M_\odot$, and the $6.5 \times 10^6 M_\odot$ MBHs in M87 (REF.122). With time, we have
accumulated evidence for a variety of empirical relations such as those between MBH mass and galaxy bulge luminosity, mass and stellar velocity dispersion\textsuperscript{[123–126]}. Massive present-day galaxies assembled through accretion from cosmic filaments and galaxy mergers, and MBHs through accretion of gas in the galaxy and MBH mergers: but what sets the relative growth of one with respect to the other? Several aspects of galaxy evolution and of MBH assembly can shape such relations (and their scatter), making their interpretation a subject of intense debate.

At the massive end of the MBH spectrum, MBHs could affect the mass growth of their host galaxies and regulate themselves through AGN feedback, whose importance is expected to decrease towards low masses. In this regime, galaxies could instead influence, potentially stunt, the growth of their MBHs through, for example, supernova feedback (FIG. 1). Therefore, such scaling relations and their dispersion could be evidence for the complex coevolution of MBHs and galaxies\textsuperscript{[127]}, and help to constrain MBH formation. Whether and how these scaling relations evolve towards higher redshifts are key questions, but selection effects and biases in observational studies make such investigations very challenging.

**Diagnostics on the origins**

The future of MBH studies is bright, as planned and proposed facilities will target the origins of MBHs using light and GW as messengers\textsuperscript{[24]}. We summarize the diagnostics in BOX 2. Seed formation channels are not mutually exclusive. Tests based on population studies can therefore identify the dominant channel, whereas tests based on individual ‘special’ outlier sources can support one particular theory without necessarily ruling out the others. To break degeneracies, one needs to combine observations of nearby galaxies with those of the farthest AGN. Traditional EM observations can give constraints on the broad population of accreting sources, through a combination of MBH mass and accretion rate, or on a smaller number of quiescent MBHs. GW observations yield directly the MBH masses, together with the distance, of merging MBHs. The key to unveiling the majority of MBHs is the synergy between these different approaches. For instance, detecting the GW signal from a binary of $\sim 100 \, M_\odot$ at $z = 10–20$ would prove that massive stars form at cosmological dawn, but whether these are seeds requires additional proof, such as detection of an accreting black hole of $10^7 \, M_\odot$ at $z > 10$. This would mean witnessing the growth of a light seed. Likewise, the GW detection of a $10^9 \, M_\odot$ MBH at redshift $z > 15$ would hint at direct collapse or early Universe cosmology channels, but if only one detects merging MBHs of $\geq 10^4 \, M_\odot$, it does not rule out the existence of light seeds. Inefficient sinking of any type of seed to the galaxy centre could be responsible for the absence of GW detections. In that case, EM emissions would be the only probes. However, if one mostly detects GWs from merging binaries of $< 10^3 \, M_\odot$, identification of heavier accreting MBHs at, say, $z > 10$ with EM telescopes would be required for heavy seed mechanisms to remain a possibility.

**Local Universe**

**EM observations.** One possibility, albeit indirect, to obtain information on the origins of MBHs is through ‘archaeology’ in local dwarf galaxies with stellar mass below $\sim 10^9 \, M_\odot$. For this, theoretical predictions need to link the ‘initial conditions’ to what we can observe today: how mass grows through accretion and mergers and the complex dynamics of sinking and ejections become key. Many studies of seeds are instead devoted only to the formation phase and do not evolve a population all the way to today, and even when they do they often include only one type of seed.

The expectation is that measuring the masses of MBHs and their occupation fraction in the smallest dwarf galaxies are ways to infer which MBH formation channel is the most common\textsuperscript{[129,130]}. Many different seed models predict similar initial masses, $\sim 10^4 \, M_\odot$ (see the figure in BOX 1), limiting the value of this probe; but the discovery of a $\sim 10^5 \, M_\odot$ MBH in a dwarf galaxy with stellar mass $\sim 10^9 \, M_\odot$ — one that can be proved not to be the stripped core of an originally larger galaxy — would hint at the existence of direct collapse.

Furthermore, MBH formation mechanisms predict that only a fraction of dwarf galaxies host MBHs, and this fraction depends on the specific mechanism. Finding MBHs in most dwarf galaxies, and especially those with stellar mass as low as $\sim 10^6 \, M_\odot$, would be a clear indication that MBH formation is common, and therefore models that predict rare seeds, such as direct collapse or some flavours of primordial black holes, can not be the dominant channel, as shown in the figure in BOX 1. This is reflected in the shape and normalization of the MBH mass distribution, known as mass function\textsuperscript{[131]}; if seeds are heavy, the mass function would be truncated at higher masses than if seeds are light. The number density also reflects the frequency, or rarity of seeds; to explain today’s MBH we need models that predict at least the observed number density.

Furthermore, finding MBHs in dwarf galaxies is observationally challenging: in galaxies with a shallow stellar distribution, MBHs may end up not being located in the galaxy centre\textsuperscript{[131]}, making their detection more difficult. Success in finding MBHs in dwarf galaxies has been obtained when selecting galaxies with dense stellar nuclei\textsuperscript{[1]}: the high stellar density makes it easier to find a MBH in the centre and to measure its mass through its dynamical signature on the stellar orbits.

Relying on AGN in dwarf galaxies is hampered by the intrinsic faintness: since MBH mass decreases with galaxy mass, dwarf galaxies host low-mass MBHs, which when accreting are faint AGN\textsuperscript{[132]}. This makes them hard to detect and disentangle from other sources that could have similar emission properties, typically binaries formed by a compact object and a star (X-ray binaries), or remnants of supernova explosions. To obtain the MBH occupation fraction, a sample based on faint AGN must also be corrected to include the MBHs that are quiescent and therefore missed by observations\textsuperscript{[135]}.

Tidal disruption events are a promising route to connect MBHs to their origins, being sensitive to the occupation fraction\textsuperscript{[134]}. Tidal disruption events occurring around MBHs lighter than about $10^6 \, M_\odot$ generally trigger
accretion episodes around or even above the Eddington luminosity, thus making these MBHs as bright as they can be for a few months, until their luminosity decays. Surveys with capabilities to detect transient sources, such as the Zwicky Transient Facility, the Legacy Survey of Space and Time (LSST) with the Rubin Telescope and eRosita will greatly increase the number of known tidal disruption events, which can then be followed up to estimate the MBH and galaxy properties, and enlarge the census of MBHs in dwarf galaxies to constrain the occupation fraction and the asymptotic MBH mass in dwarfs. The MBH masses can be estimated from the width of emission lines, using virial arguments, in the case of active MBHs or measured through the kinematics of stars and gas, which have been proven successful even in the dwarf galaxy regime.

Discovering black holes of ~10^4 M⊙ in the star clusters of nearby (star-forming) galaxies would provide compelling evidence that dynamical interactions, runaway or hierarchical, leading to black holes of 10^2−10^3 M⊙ occur in nature. Metallicity, age and compactness of the cluster could help refine which of the two avenues, runaway versus hierarchical merger, is the most likely.

**GW observations.** The recent discovery of the binary that produced the heaviest stellar-mass black hole yet observed of 142 ± 28 M⊙ has already sparked a lively discussion on its origin — whether it formed in a field binary, or via stellar collisions, or by hierarchical mergers of lower-mass black holes in star cluster, that is, by the same mechanisms proposed for the formation of seeds. Although the GW antennas on Earth are already uncovering systems that could represent the low-redshift analogues of seeds, space-borne LISA can contribute to our understanding of the origins of MBHs with extreme-mass-ratio inspirals (EMRIs). EMRIs occur when a compact object, such as a stellar-mass black hole, is captured on a relativistic orbit by an MBH and slowly inspirals solely by emission of GWs, covering 10^4−5 orbits until it crosses the horizon. EMRIs are of pivotal importance to map the spacetime and test the nature of gravity, and the MBH mass and spin can be measured with exquisite precision. LISA can detect EMRIs around MBHs with mass ~10^4−10^5 M⊙, with systems of mass ~10^4 M⊙ detectable up to z ~ 4 (Ref. 138). If LISA discovers enough EMRIs, one can probe the MBH mass function at low mass, which, as discussed in the previous section, is a sensitive probe of seed formation channels. The combination of the information from EMRIs and EM searches in dwarf galaxies represents an excellent synergy to constrain seed formation. The rate of EMRIs is currently unconstrained: different mechanisms, each with large uncertainties, can lead to their formation, from the classic mechanism of stellar dynamical relaxation around an MBH, to breaking of binaries composed of a stellar black hole and a star by the tidal field of the MBH, to migration in the accretion disks of AGN.

**High-redshift Universe**

**EM observations.** Beyond z ≥ 6, our current facilities only allow us to detect the EM signal of the most massive, M_{BH} ≥ 10^6 M⊙, and luminous quasars with bolometric luminosity of L_{bol} ≥ 10^{46} erg s^{-1}. These rare objects increased their mass by many orders of magnitude since birth, erasing the memory of the seeding and growth history. Thus, they provide us with degenerate constraints on their initial seed mass, and in addition, they represent only the tip of the iceberg of the whole MBH population. Addressing the origins of MBHs requires to be able to find seed-mass MBHs close to the redshifts of their formation. Hard X-rays can penetrate through the dense gas clouds in the vicinity of the MBH, whereas optical radiation cannot. Thus, next-generation X-ray missions such as Athena, or concepts as Lynx and AXIS, aim at a substantial increase in sensitivity, in order to be able to detect young accreting MBHs. Detecting and measuring the luminosity, and counting MBHs to build a luminosity function, is not enough, though — this will only constrain a combination of MBH parameters: seed mass distribution, accretion distribution, and probability for galaxies to host an MBH.

The synergy between these X-ray missions and the near-future deep imaging surveys of the James Webb Space Telescope (JWST, near/mid-infrared) and Nancy Grace Roman Space Telescope (optical–near-infrared) will help in breaking these degeneracies. Different MBH formation mechanisms yield a different connection between the MBHs and their host galaxies. Small M_{BH}/M_{⋆} mass ratios are favoured by mechanisms forming seeds as the remnants of Population III stars, as runaway stellar mergers in compact nuclear clusters, or as runaway black hole mergers. In the direct-collapse model, instead, newly formed MBHs are predicted to be overmassive compared with their host galaxies, with large M_{BH}/M_{⋆} ratios of about unity. Therefore, the combination of MBH and galaxy properties could provide us with direct high-redshift constraints on MBH origins. The spectral energy distribution of high-redshift systems, observed with JWST, could also help us to distinguish between different seeding mechanisms. Although in the case of light seeds the MBH emission would be outshone by the stellar component of the host galaxies and hence missed, direct-collapse MBHs could outshine their galaxies in the infrared and would be detectable. Finally, within a narrow mass range, SMSs could explode rather than collapse into black holes: by extrapolation, detection of such hypernovae could be a proof of the existence of SMSs outside this interval.

**GW observations.** A network of next-generation terrestrial interferometers (3G), and LISA or similar interferometers from space, will probe an immense volume of the Universe and of the MBH parameter space as their horizons extend up to z ~ 20−30 and even beyond. This will enable us probe black hole masses from 100 M⊙ up to 10^7 M⊙. 3G detectors with sensitivity from a few Hz to a few kHz have the unique capability of discovering the earliest binary light seeds forming in the Universe, and also those that failed to grow, the starved seeds that might be still present in galaxies at lower redshift. LISA, as an all-sky monitor sensitive to frequencies between 10^{-4} and 10^{-1} Hz, has the potential to detect...
Penrose’s concept of black hole formation

Uncovering the origins of massive black holes (MBHs) is an intellectual voyage into one of the most fascinating discoveries of general relativity: the existence of absolute event horizons, responsible for the blackness of the ‘holes’ emerging when gravitational collapse passes a point of no return. Roger Penrose introduced powerful concepts: the idea of a trapped surface, formalizing the condition of light-ray convergence in collapsing matter, and of singularity of spacetime, which he characterized through the notion of incomplete causal geodesics. In his endless effort to understand the causal and global structure of spacetime, Penrose postulated the idea of ‘weak cosmic censorship’ stating that singularities are inaccessible to external observers, covered by the event horizon.

Penrose gave an illuminating description of gravitational collapse to a black hole\(^1\):
“A body, or collection of bodies, collapses down to a size comparable to its Schwarzschild radius, after which a trapped surface can be found in the region surrounding the matter. Some way outside the trapped surface region is a surface which will ultimately be the absolute event horizon. But at present, this surface is still expanding somewhat. Its exact location is a complicated affair and it depends on how much more matter (or radiation) ultimately falls in. We assume only a finite amount falls. Then the expansion of the absolute event horizon gradually slows down to stationarity. Ultimately the field settles down to becoming a Kerr solution.”

Nature seems to have brought all of this into physical reality. This is testified by the discovery, through precision stellar-dynamical measurements by Reinhard Genzel and Andrea Ghez, of a massive dark object in the midst of our Milky Way Galaxy, for which the MBH hypothesis appears the most natural, physical explanation. The Event Horizon Telescope has also provided a tantalizing image of light close to the event horizon of the MBH in M87 (REF\(^{36}\)).

Outlook

Black holes have the beauty of being the simplest, most elementary astrophysical objects in the Universe. They are, however, involved in complex physical processes once they interact with their environment. They are powerful engines, being bright sources in the EM and GW skies. Our astrophysical understanding of black holes is rooted in the recognition that relatively massive stars collapse into neutron stars, and that the massive ones collapse into stellar black holes, since there is an upper bound to the mass of a neutron star. The physics leading to the formation of stellar black holes stands on solid ground, despite uncertainties on the state of matter above nuclear densities. By contrast, the physics leading to the formation of MBHs is not known. MBHs may originate from primordial black holes, or from relatively light black holes, both rearranged and transformed into giants by accretion and mergers. SMBs would die as heavy seeds in response to general relativity instabilities excited inside radiation-dominated, low-density plasmas whose microphysical state is well known. Alas, we do not have observational evidence yet of the collapse ‘to the point of no return’ of such massive stars (BOX 4).

What we do know is that upcoming ground-based or space-based telescopes aim at constraining the AGN population at high redshift, potentially down to low-mass MBHs close to the seeding mass, whereas the detections of MBHs in local dwarf galaxies can provide indirect constraints by analogy with high-redshift systems. These EM observations complement GWs that will constrain the merging MBH population (\(10^2–10^3 \text{M}_\odot\)), active or quiescent, but are also limited to the binary population. We are entering the era of ‘precision GW astrophysics,’ detecting signals from coalescing black holes on all mass scales to witness the emergence of astrophysical event horizons in dynamical spacetimes, and signals from EMRIs, the sources that will let us probe the global, causal structure of stationary spacetimes around central dark massive objects.

There are various facets to the origins of MBHs, and we have to consider them together since they are interrelated. We need to be open-minded about connecting different fields of research — don’t just look under the lamppost, there is another lamppost not far away.

Published online 20 September 2021

---

1. Nguyen, D. D. et al. Improved dynamical constraints on the masses of the central black holes in nearby low-mass early-type galactic nuclei and the first black hole determination for NGC 205. *Astrophys. J.*, **782**, 104 (2019).
2. Abbott, R. et al. Properties and astrophysical implications of the 150 M\(_\odot\) binary black hole merger GW190521. *Astrophys. J. Lett.*, **900**, L13 (2020).
3. Fabian, A. C. Observational evidence of active galactic nuclei feedback. *Ann. Rev. Astron. Astrophys.*, **50**, 455–489 (2012).
4. Fan, X. et al. A survey of z>5.8 quasars in the Sloan digital sky survey. I. Discovery of three new quasars and the spatial density of luminous quasars at z ~ 6. *Astron. J.*, **122**, 2835–2849 (2001).
5. Yu, Q. & Tremaine, S. Observational constraints on growth of massive black holes. *Mon. Not. R. Astron. Soc.*, **355**, 965–976 (2002).
6. Mezcua, M. Observational evidence for intermediate-mass black holes. *Int. J. Mod. Phys. D*, **26**, 1750021 (2017).
7. Greene, J. E., Strader, J. & Ho, L. C. Intermediate-mass black holes. *Annu. Rev. Astron. Astrophys.*, **58**, 257–312 (2020).
8. Rees, M. J. in Structure and Properties of Nearby Galaxies vol. 77 (eds Berkhuijsen, E. M. & Weibelbrink, R.) 237–242 (Springer, 1978).
9. Inayoshi, K., Volbul, E. & Haiman, Z. The assembly of the first massive black holes. *Annu. Rev. Astrophys.*, **58**, 77–97 (2020).
10. Palla, F., Salpeter, E. E. & Stahler, S. W. Primordial star formation — the role of molecular hydrogen. *Astrophys. J.*, **271**, 625–641 (1986).
11. Vink, J. S., de Koter, A. & Lamers, H. G. J. L. M. Mass-loss predictions for O and B stars as a function of metallicity. *Astrophys. J.*, **569**, 574–588 (2002).
12. Spera, M., Magelli, M. S. & Bressan, A. The mass spectrum of compact remnants from the PARSEC stellar evolution tracks. *Mon. Not. R. Astron. Soc.*, **451**, 4086–4105 (2015).
13. Madau, P. & Rees, M. J. Massive black holes as population III remnants. *Astrophys. J. Lett.*, **551**, L27–L30 (2001).
14. Abel, T., Bryan, G. L. & Norman, M. L. The formation of the first star in the Universe. *Science*, **295**, 55–98 (2002).
15. Greif, T. H. et al. Simulations on a moving mesh: the clustered formation of Population III protostars. *Astrophys. J.*, **737**, 75 (2011).
16. Hosokawa, T. et al. Formation of massive primordial stars: intermittent UV feedback with episodic mass accretion. *Astrophys. J.*, **824**, 119 (2016).
17. Schneider, R., Ferrara, A., Natarajan, P. & Omukai, K. First stars, very massive black holes, and metals. *Astrophys. J.*, **571**, 50–99 (2002).
18. Tarumi, Y., Hartwig, T. & Magg, M. Implications of inhomogeneous metal mixing for stellar archaeology. *Astrophys. J.*, **897**, 58 (2020).
19. Bromm, V. & Loeb, A. Formation of the first supermassive black holes. *Astrophys. J.*, **596**, 34–66 (2003).
20. Begelman, M. C., Volonteri, M. & Rees, M. J. Formation of supermassive black holes by direct collapse in pre-galactic haloes. *Mon. Not. R. Astron. Soc.*, **370**, 289–298 (2006).
Habouzit, M., Volonteri, M. & Dubois, Y. Blossoms from black hole seeds: properties and early growth regulated by supernova feedback. Mon. Not. R. Astron. Soc. 468, 3935–3948 (2017).

Bower, R. G. et al. The first seeds of galaxy formation: why haloes trigger black hole growth and bring star formation to an end. Mon. Not. R. Astron. Soc. 465, 12–18 (2017).

Belloyer, J. M. et al. Multimessenger signatures of massive black holes in dwarf galaxies. Mon. Not. R. Astron. Soc. 486, 1011–1019 (2019).

Sijacki, D., Springel, V. & Haehnelt, M. Growing black holes in nuclear star clusters with gas inflows. Mon. Not. R. Astron. Soc. 442, 3616–3626 (2014).

Vavilov, E., Haiman, Z. & Yajima, H. & Khochfar, S. The role of stellar relaxation in rapidly growing pre-galactic gas clouds. Mon. Not. R. Astron. Soc. 486, 5008–5021 (2019).

Davies, M. B., Miller, C. G. & Belloyer, J. M. Black hole supermassive formation via gas accretion in nuclear stellar clusters. Astrophys. J. Lett. 740, L42 (2014).

Lupi, A., Colpi, M., Devecchi, B., Galanti, G. & Volonteri, M. Multigeneration black hole formation in black hole seeds in nuclear star clusters with gas inflows. Mon. Not. R. Astron. Soc. 442, 3616–3626 (2014).

Alexander, T. & Natarajan, P. Rapid growth of seed black holes in the early Universe by super-exponential accretion. Science 345, 1330–1334 (2014).

Arcavi, S. M., Bellovary, J. M. & Peirani, S. On the number density of ‘direct collapse’ black hole formation sites hostile to atomic cooling haloes. Mon. Not. R. Astron. Soc. 486, 5008–5021 (2019).

Davies, M. B., Miller, C. G. & Belloyer, J. M. Black hole supermassive formation via gas accretion in nuclear stellar clusters. Astrophys. J. Lett. 740, L42 (2014).

Lupi, A., Colpi, M., Devecchi, B., Galanti, G. & Volonteri, M. Multigeneration black hole formation in black hole seeds in nuclear star clusters with gas inflows. Mon. Not. R. Astron. Soc. 442, 3616–3626 (2014).

Garcia-Belido, J. Primordial black hole seeds and the origin of the matter–antimatter asymmetry. Phil. Trans. R. Soc. Lond. A 377, 20190099 (2019).

Rubin, S. N., Sakharov, A. S. & Khlopov, M. Y. The formation of primary galactic nuclei during phase transitions in the early universe. Sov. J. Nucl. Phys. 17, 122–127 (1973).

Garcia-Belido, J. Primordial black hole seeds and the origin of the matter–antimatter asymmetry. Phil. Trans. R. Soc. Lond. A 377, 20190099 (2019).

Rubin, S. N., Sakharov, A. S. & Khlopov, M. Y. The formation of primary galactic nuclei during phase transitions in the early universe. Sov. J. Nucl. Phys. 17, 122–127 (1973).

Garcia-Belido, J. Primordial black hole seeds and the origin of the matter–antimatter asymmetry. Phil. Trans. R. Soc. Lond. A 377, 20190099 (2019).

Rubin, S. N., Sakharov, A. S. & Khlopov, M. Y. The formation of primary galactic nuclei during phase transitions in the early universe. Sov. J. Nucl. Phys. 17, 122–127 (1973).

Garcia-Belido, J. Primordial black hole seeds and the origin of the matter–antimatter asymmetry. Phil. Trans. R. Soc. Lond. A 377, 20190099 (2019).

Rubin, S. N., Sakharov, A. S. & Khlopov, M. Y. The formation of primary galactic nuclei during phase transitions in the early universe. Sov. J. Nucl. Phys. 17, 122–127 (1973).

Garcia-Belido, J. Primordial black hole seeds and the origin of the matter–antimatter asymmetry. Phil. Trans. R. Soc. Lond. A 377, 20190099 (2019).

Rubin, S. N., Sakharov, A. S. & Khlopov, M. Y. The formation of primary galactic nuclei during phase transitions in the early universe. Sov. J. Nucl. Phys. 17, 122–127 (1973).
